

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compen Ltd.

OM protein - protein search, using sw model

Run on: February 19, 2004, 10:01:13 ; Search time 41 Seconds
(without alignments)
42.585 Million cell updates/sec

Title: US-09-551-336b-1
Perfect score: 45
Sequence: 1 AAAAAAAAAAK 11

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : A.Geneseq.15Jun03.*
1: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA2001.DAT:*
23: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA2002.DAT:*
24: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
45	100.0	16	18	AAW32224	Beta-sheet forming
45	100.0	16	18	AAW32226	Beta-sheet forming
45	100.0	17	18	AAW32227	Beta-sheet forming
45	100.0	20	24	ABP59409	Self-assembling di
45	100.0	30	24	ABP59422	Self-assembling tr
45	100.0	30	24	ABP59423	Self-assembling tr
45	100.0	97	22	ABG75920	Human colon cancer
45	100.0	201	23	ABP41465	Human ovarian anti
45	100.0	216	23	ABP62982	Human polypeptide

10	45	100.0	218	21	AAW00755
11	45	100.0	218	21	AAW00759
12	45	100.0	220	23	AAU76972
13	45	100.0	220	24	ABU07443
14	45	100.0	221	23	ABP62855
15	45	100.0	225	21	AAW00758
16	45	100.0	226	17	AAW05151
17	45	100.0	262	22	ABG15586
18	45	100.0	265	21	ABP58221
19	45	100.0	335	22	ABP58955
20	45	100.0	406	22	ABP58251
21	45	100.0	515	21	AAW6135
22	45	100.0	571	21	AAW69071
23	45	100.0	660	20	AAW01303
24	45	100.0	698	20	AAW01302
25	45	100.0	698	21	AAW69069
26	45	100.0	712	21	AAW08630
27	45	100.0	730	19	AAW46315
28	45	100.0	730	21	AAW08631
29	45	100.0	730	23	AAW01360
30	45	100.0	731	21	AAW69068
31	45	100.0	731	22	ABP6657
32	45	100.0	733	15	AAW5653
33	45	100.0	733	20	AAW01301
34	42	93.3	20	24	ABP59411
35	42	93.3	30	24	ABP59424
36	42	93.3	30	24	ABP59425
37	42	93.3	31	21	AAW08166
38	42	93.3	102	24	ABU20243
39	42	93.3	102	24	ABU20253
40	42	93.3	102	24	ABU20256
41	42	93.3	157	22	AAU16372
42	42	93.3	157	24	ABU5441
43	42	93.3	161	21	AAW19319
44	42	93.3	229	23	ABP41248
45	42	93.3	244	22	ABG21054

ALIGNMENTS

RESULT 1
ID AAW32224 strand: peptide; 16 AA.

AC AAW32224;

DT 12-FEB-1998 (first entry)

DE Beta-sheet forming peptide #1.

KW Beta sheet; circular dichroism spectroscopy; glycosidase activity;

KW phospholipase activity; drug screening; hydrolysis;

KW neurodegenerative disease; Alzheimer's disease; amyloid deposition.

OS Synthetic.

FT Key Location/Qualifiers

FT Modified-site 1 /note="N-terminal acetyl"

FT Modified-site 16 /note="C-terminal amide"

PN W09637212.A1.

PD 28-NOV-1996.

PF 23-MAY-1996; 96WO-US07564.

PR 26-MAY-1995; 95US-0452043.

PA (TORR-) TORREY PINES INST MOLECULAR STUDIES.

Human secreted pro
Human secreted pro
Human ribosomal L1
Protein different
Human polypeptide
Human secreted pro
Nucleic proliferat
Novel human diagno
Lung cancer associ
Drosophila melanog
Human protein sequ
Amino acid sequenc
Human tropoelastin
Human tropoelastin
Amino acid sequenc
Amino acid sequenc
Human elastin prot
Synthetic human tr
Amino acid sequenc
Self-assembling di
Self-assembling tr
Peptide modulating
162P16 cancer gen
162P16 cancer gen
Human novel secret
Human novel polype
Amino acid sequenc
Human ovarian anti
Novel human diagno

WPI; 1997-033944/03.

Polypeptide(s) which form stable beta sheets in aq. environment - having phosphodiesterase and glycosidase activity and useful in bioengineering, enzymatic and drug screening applications

Example 2; Page 16; 35pp; English.

This sequence represents a specific example of a polypeptide, which forms a beta sheet in an aqueous environment and has the general formula JnAmXAm2n or Ac-JnAmXAm2n-NH_2 , in which A = D- or L-alanine or amino acid; m1, m2 = 0-40; m1+m2 = 10-40; J = a charged amino acid; n = 1 or 2; X = any amino acid except proline; u = 0 or 1; yl lysine, gamma-oxyllysine, arginine, histidine, aspartic acid, glutamic acid or gamma-carboxyglutamic acid; Ac = an acylation modification to the amino terminus; NH2 = an amidation modification to the carboxylic acid terminus; and X1 = cysteine, threonine, tyrosine or serine. The peptides have phosphodiesterase activity and readily hydrolyse nucleic acids. They also have glycosidase activity and readily hydrolyse sugar phosphates. Further they can cause amine-catalysed decarboxylation (e.g. decarboxylation of oxalacetate), and they can hydrolyse phospholipids. They can be used for bioengineering, enzymatic and drug screening applications. In particular they can be used for screening drugs for the prevention or treatment of neurodegenerative diseases, e.g. Alzheimer's disease, which result in amyloid protein deposition. They can also be used to screen for drugs which inhibit or disrupt the beta sheet.

Sequence 17 AA;

Query Match 100.0%; Score 45; DB 18; Length 17;
Best Local Similarity 100.0%; Pred. No. 0.33;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 AAAAAAAAAAK 11
|||||||
6 AAAAAAAAAAK 16

RESULT 4
ABP59409

ID ABP59409 standard; peptide; 20 AA.

ABP59409;

09-JUN-2003 (first entry)

Self-assembling di-block oligopeptide AK20.

Self-assembling oligopeptide; dipolar; nanotube; nanostructure;
drug delivery; carrier.

Synthetic.

WO2003006043-A1.

23-JAN-2003.

10-JUL-2002; 2002WO-US21757.

10-JUL-2001; 2001US-304256P.

(MASI) MASSACHUSETTS INST TECHNOLOGY.

Zhang S, Vauthey S;

WPI; 2003-267944/26.

New dipolar oligopeptides in a self-assembled nanostructure useful for drug delivery

Claim 16; Page 21; 52pp; English.

The present invention relates to self-assembling dipolar oligopeptides

and di- and tri-block peptide copolymers. The dipolar oligopeptides have ability to self assemble to form stable nanotubes. The self-assembled nanostructure have the ability to entrap and deliver molecules with high degree of efficacy. The copolymers are amenable for molecular systematic design, modelling and simulations before synthesis; can be highly purified to be mono-dispersed materials; combinatorial approach can be employed to systematically characterize these co-polymers at various ratio; and can be synthesized in vitro or in vivo. The nanotubes fuse with the lipid bilayers and do not deform the cells as liposomes do. The oligopeptides can be used in a self-assembled nanostructure (e.g. nanotube) for delivering drug into a cell and as a carrier for biologically active materials. The present sequence is one such di-block oligopeptide.

Sequence 20 AA;

Query Match 100.0%; Score 45; DB 24; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.38;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 AAAAAAAAAAK 11
|||||||
1 AAAAAAAAAAK 11

RESULT 5
ABP59422

ID ABP59422 standard; peptide; 30 AA.

ABP59422;

09-JUN-2003 (first entry)

Self-assembling tri-block oligopeptide KAK30.

Self-assembling oligopeptide; dipolar; nanotube; nanostructure;
drug delivery; carrier.

Synthetic.

WO2003006043-A1.

23-JAN-2003.

10-JUL-2002; 2002WO-US21757.

10-JUL-2001; 2001US-304256P.

(MASI) MASSACHUSETTS INST TECHNOLOGY.

Zhang S, Vauthey S;

WPI; 2003-267944/26.

New dipolar oligopeptides in a self-assembled nanostructure useful for drug delivery

Claim 21; Page 23; 52pp; English.

The present invention relates to self-assembling dipolar oligopeptides and di- and tri-block peptide copolymers. The dipolar oligopeptides have ability to self assemble to form stable nanotubes. The self-assembled nanostructure have the ability to entrap and deliver molecules with high degree of efficacy. The copolymers are amenable for molecular systematic design, modification and synthesis; can be subjected to extensive molecular modelling and simulations before synthesis; can be highly purified to be mono-dispersed materials; combinatorial approach can be employed to systematically characterize these co-polymers at various ratio; and can be synthesized in vitro or in vivo. The nanotubes fuse with the lipid bilayers and do not deform the cells as liposomes do. The oligopeptides can be used in a self-assembled nanostructure (e.g. nanotube) for delivering drug into a cell and as a carrier for

CC biologically active materials. The present sequence is one such
CC tri-block oligopeptide.

XX
SQ Sequence 30 AA;

Query Match 100.0%; Score 45; DB 24; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.56;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAK 11

Db 11 AAAAAAAAAAK 21

RESULT 6

ABP59423

ID ABP59423 standard; peptide; 30 AA.

AC ABP59423;

DT 09-JUN-2003 (first entry)

DE Self-assembling tri-block oligopeptide AKA30.

XX Self-assembling oligopeptide; dipolar; nanotube; nanostructure;
KM drug delivery; carrier.

XX Synthetic.

OS WO200306043-A1.

PN 23-JAN-2003.

PD 10-JUL-2002; 2002WO-US21757.

XX 10-JUL-2001; 2001US-304256P.

XX (MASI) MASSACHUSETTS INST TECHNOLOGY.

PI Zhang S, Vauthey S;

DR WPI; 2003-267944/26.

XX New dipolar oligopeptides in a self-assembled nanostructure useful for
PT drug delivery -

PS Claim 23; Page 23; 52pp; English.

XX The present invention relates to self-assembling dipolar oligopeptides
CC and di-and tri-block peptide copolymers. The dipolar oligopeptides have
CC ability to self assemble to form stable nanotubes. The self-assembled
CC nanostructure have the ability to entrap and deliver molecules with high
CC degree of efficacy. The copolymers are amenable for molecular systematic
CC design, modification and synthesis; can be subjected to extensive
CC molecular modelling and simulations before synthesis; can be highly
CC purified to be mono-dispersed materials; combinatorial approach can be
CC employed to systematically characterize these co-polymers at various
CC ratios; and can be synthesized in vitro or in vivo. The nanotubes fuse
CC with the lipid bilayers and do not deform the cells as liposomes do.
CC The oligopeptides can be used in a self-assembled nanostructure (e.g.
CC nanotube) for delivering drug into a cell and as a carrier for
CC biologically active materials. The present sequence is one such
CC tri-block oligopeptide.

XX Sequence 30 AA;

Query Match 100.0%; Score 45; DB 24; Length 30;

Best Local Similarity 100.0%; Pred. No. 0.56;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAK 11

Db 1 AAAAAAAAAAK 11

RESULT 7

AA675920

ID AA675920 standard; Protein; 97 AA.

XX AA675920;

DT 03-SEP-2001 (first entry)

DE Human colon cancer antigen protein SEQ ID NO:6684.

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
KM colorectal carcinoma.

XX Homo sapiens.

OS WO200122920-A2.

PN 05-APR-2001.

PD 28-SEP-2000; 2000WO-US26524.

XX 28-SEP-1999; 99US-0157137.

PR 03-NOV-1999; 99US-0163280.

XX (HUMA-) HUMAN GENOME SCI INC.

PI Ruben SM, Barash SC, Birse CE, Rosen CA;

DR WPI; 2001-235357/24.

XX N-PSDB; AAH35325.

PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers -

PS Claim 11; Page 8150; 9803pp; English.

XX AAH32943 to AAH37195 and AA673514 to AA677788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patient's own production of P.
CC Additionally, N may be used to produce the colon cancer-associated P,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AA677789 represent sequences used in the exemplification of the
CC present invention.

CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX Sequence 97 AA;

Query Match 100.0%; Score 45; DB 22; Length 97;

Best Local Similarity 100.0%; Pred. No. 1.7;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAK 11

Db 77 AAAAAAAAAAK 87

RESULT 8

ABP41465

ID ABP41465 standard; Protein; 201 AA.

AC ABP41465;
 XX 22-AUG-2002 (first entry)
 XX
 DE Human ovarian antigen HOOJ038, SEQ ID NO:2597.
 KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
 KW ovarian cancer; breast cancer; tumour; reproductive system disorder;
 KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
 KW PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;
 KW inflammatory condition; immune disorder; blood disorder;
 KW cardiovascular disorder; respiratory disorder; neurological disorder;
 KW gastrointestinal disorder; urinary system disorder; drug screening;
 KW gene therapy; chromosome mapping; forensic analysis;
 KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;
 KW antiinflammatory; gynaecological; reproductive.
 XX
 OS Homo sapiens.
 XX
 PN WO200200677-A1.
 XX
 PD 03-JAN-2002.
 XX
 PF 07-JUN-2001; 2001WO-US18569.
 XX
 PR 07-JUN-2000; 2000US-209467P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Birse CE, Rosen CA;
 PT WPI: 2002-147878/19.
 DR N-PSDB; ABQ54542.
 XX
 PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,
 PT useful in the prevention, treatment and diagnosis of cancer (e.g.
 PT ovarian cancer), immune disorders, cardiovascular disorders and
 PT neurological diseases -
 XX
 PS Claim 11; SEQ ID NO 2597; 2922pp; English.

The invention relates to 2175 novel human ovarian antigens (ABP41054-
 CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also
 CC encompasses polypeptides 90% identical and polynucleotides 95% identical
 CC to the sequences of the invention. The invention additionally relates to
 CC recombinant vectors and host cells comprising human ovarian antigen
 CC polynucleotides, antibodies against human ovarian antigens, and the use
 CC of ovarian antigen polynucleotides and polypeptides in diagnosing,
 CC treating, prognosing or preventing various ovary and/or breast-related
 CC disorders. Such conditions include ovarian cancer and breast cancer, and
 CC metastatic tumours of ovarian or breast origin, reproductive system
 CC disorders (e.g., infertility, disorders of pregnancy, anovulation,
 CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhea), endocrine
 CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
 CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
 CC vaginitis), immune disorders (e.g., congenital and acquired
 CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
 CC blood-related disorders (e.g., anaemia), cardiovascular disorders,
 CC respiratory disorders, neurological disorders, gastrointestinal disorders
 CC and urinary system disorders. Ovarian antigen polypeptides and
 CC polynucleotides may also be used in screening for compounds which
 CC modulate ovarian antigen expression or activity. The polynucleotides may
 CC further be used for gene therapy, chromosome mapping, in the
 CC identification of individuals and in forensic analysis, and the
 CC polypeptides may be used as food additives or to prepare antibodies
 CC useful in disease diagnosis, drug targeting and phenotyping. The present
 CC sequence in represents a human ovarian antigen of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

Sequence 201 AA;

Query Match 100.0%; Score 45; DB 23; Length 201;
 Best Local Similarity 100.0%; Pred. No. 3.3;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAK 11
 DB 136 AAAAAAAAAAK 146

RESULT 9
 ID ABP62982 standard; Protein; 216 AA.
 AC ABP62982;
 XX
 DT 14-OCT-2002 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 419.
 XX
 KW Human; vulnery; dermatological; neuroprotective; nootropic; cancer;
 KW antiparkinsonian; immunostimulant; cytostatic; immunosuppressive;
 KW antidiabetic; antiallergic; gene therapy; wound healing; tissue repair;
 KW burn; central nervous system disorder; Alzheimer's disease;
 KW Parkinson's disease; Huntington's disease; immune disorder;
 KW autoimmune disorder; multiple sclerosis; diabetes; allergy.
 XX
 OS Homo sapiens.
 XX
 PN WO200218424-A2.
 XX
 PD 07-MAR-2002.
 XX
 PF 31-AUG-2001; 2001WO-US27093.
 XX
 PR 01-SEP-2000; 2000US-0654935.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Asundi V, Zhou P, Xue AJ, Ren F, Zhang J, Wang J;
 PI Zhao QA, Wang D, Liu C, Drmanac RT, Wehman T;
 XX
 DR WPI: 2002-583321/62.
 XX
 DR N-PSDB; ABQ93461.
 XX
 PT New polynucleotide and polypeptides, useful for treatment and diagnosis
 PT of Alzheimer's, Parkinson's, Huntington's, amyotrophic lateral
 PT sclerosis, immune deficiencies, cancer, autoimmune disorders, multiple
 PT sclerosis, diabetes and allergies -
 XX
 PS Claim 20; SEQ ID NO 419; 284pp + Sequence Listing; English.

The invention relates to an isolated polynucleotide (I) comprising one of
 CC 245 sequences (ABQ93288-ABQ93532). Treating a condition comprising
 CC administering to a mammalian subject a composition comprising the protein
 CC (II) encoded by (I) (ABP62809-ABP63053) or an antibody (III) to (II).
 CC (I), (II) and (III) are useful for diagnostic evaluation of disorders.
 CC (I) is useful for gene therapy of diseases and (II) can be used for
 CC therapeutic treatment. Diseases that may be treated include wound healing
 CC and tissue repair, burns, central nervous system disorders (e.g.
 CC Alzheimer's, Parkinson's, Huntington's and amyotrophic lateral
 CC sclerosis), immune deficiencies, cancer, autoimmune disorders, multiple
 CC sclerosis, diabetes and allergies.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

Sequence 216 AA;

Query Match 100.0%; Score 45; DB 23; Length 216;
 Best Local Similarity 100.0%; Pred. No. 3.6;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAK 11

XX 14-FEB-2002.
PD 06-AUG-2001; 2001WO-US24718.
XX 04-AUG-2000; 2000US-222811P.
XX (TEXA) UNIV TEXAS SYSTEM.
PA Katz R, Jiang F;
XX WPI; 2002-217200/27.
DR N-PSDB; ABR10349.
PT Identifying subject at risk for development of cancer, preferably lung
PT cancer, comprises contacting RPL14, CD39L3, PMGM, or GC20 gene probe
PT with test sample obtained from subject, and analysing DNA from test
PT sample -
XX Example 1; Page 69-70; 79pp; English.
XX The invention describes a method of identifying a subject at risk for the
CC development of cancer, predicting progression or metastasis of non-small
CC cell carcinoma and other carcinoma in a subject, or identifying an
CC individual to be segregated from a high risk environment. The method
CC comprises contacting an RPL14, CD39L3, PMGM, or GC20 gene probe with a
CC test sample obtained from a subject, and analysing DNA from the test
CC sample. The method is useful for identifying a subject (a smoker,
CC non-smoker or former smoker) at risk for the development, recurrence, or
CC metastasis of cancer (preferably cancer of lung, bladder, head, neck,
CC urothelial, kidney, pancreas, mouth, throat, pharynx, larynx or
CC esophagus, or an upper airway primary or secondary cancer), to identify
CC subjects who need an intensive follow-up protocol and for the prognosis
CC and diagnosis of cancer. This is the amino acid sequence of the human
CC ribosomal L14 protein (RPL14, located on chromosome 3p21.3), used to
CC develop the gene probe described in the method of the invention.
SQ Sequence 220 AA;
Query Match 100.0%; Score 45; DB 23; Length 220;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAAX 11
Db 155 AAAAAAAAAAAX 165
RESULT 13
ABU07443
ID ABU07443 standard; Protein; 220 AA.
XX
AC ABU07443;
XX
DT 28-JAN-2003 (first entry)
XX
DE Protein differentially regulated in prostate cancer #46.
XX
KW Prostate cancer; gene expression; differential regulation;
KW molecular marker; drug target; cancer detection; cancer diagnosis;
KW cancer staging; cancer grading; cancer assessing; cancer monitoring.
XX
OS Homo sapiens.
XX
XX WO200281638-A2.
XX
XX 17-OCT-2002.
XX
XX 08-APR-2002; 2002WO-US10824.
XX
XX 06-APR-2001; 2001US-28171P.
XX
XX 06-APR-2001; 2001US-281732P.
XX

PA (ORIG-) ORIGENE TECHNOLOGIES INC.
XX Sun Z, Jay G;
PI WPI; 2003-058520/05.
XX
XX Novel genes which are differentially regulated in prostate cancer,
PT useful for diagnosing prostate cancer in prostate tissue sample and
PT assessing therapeutic or preventive intervention in prostate cancer
PT patients -
XX Claim 1; Page 297; 416pp; English.
XX
XX The invention describes genes (I) which are differentially regulated in
CC prostate cancer. (I) is useful for diagnosing a prostate cancer in a
CC sample comprising prostate tissue, which involves determining the number
CC of target genes which are differentially-regulated in the sample, where
CC the number is indicative of the probability that the sample comprises
CC prostate cancer. (I) is useful for assessing a therapeutic or preventive
CC intervention in a subject having a prostate cancer, which involves
CC determining the expression levels in a sample comprising prostate tissue
CC of target genes which are differentially-regulated in prostate cancer.
CC Preferably, the expression levels of at least 10 genes are determined.
CC (I) is also useful for identifying agents that modulate a biological
CC activity of a polypeptide differentially-regulated in prostate cancer
CC cells, which involves contacting a polypeptide differentially-regulated
CC in prostate cancer cells with a test agent under conditions effective for
CC the test agent to modulate a biological activity of the polypeptide, and
CC determining whether the test agent modulates the biological activity.
CC (I) is useful as molecular markers, as drug targets, and for detecting,
CC diagnosing, staging, grading, assessing, monitoring, prognosticating,
CC preventing or treating, determining predisposition to diseases and
CC conditions especially relating to prostate cancer. (I) and its expression
CC products are used in the diagnostic test to assay for presence of cancer
CC e.g., in tissue sections, in biopsy sample, in total RNA, in lymph, in
CC blood etc. (I) is useful for assessing cancer e.g., to determine the type
CC of cancer, its stage of development, the nature of genetic defect, etc.
CC The polypeptide encoded by (I) can be used as target for therapy or drug
CC discovery. (I) can also be used for expressing the polypeptide and thus
CC for searching specific binding partners of the polypeptide. (I) is
CC useful in therapeutic applications to treat prostate cancer. The
CC identification of specific genes, and groups of genes, expressed in
CC pathways physiologically relevant to prostate cancer permits the
CC definition of functional and disease pathways and the delineation of
CC targets in these pathways which are useful in diagnostic, therapeutic,
CC and clinical applications. This is the amino acid sequence of a protein
CC differentially regulated in prostate cancer.
XX
SQ Sequence 220 AA;
Query Match 100.0%; Score 45; DB 24; Length 220;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAAX 11
Db 155 AAAAAAAAAAAX 165
RESULT 14
ABP62855
ID ABP62855 standard; Protein; 221 AA.
XX
AC ABP62855;
XX
DT 14-OCT-2002 (first entry)
XX
DE Human polypeptide SEQ ID NO 292.
XX
XX Human; vulnery; dermatological; neuroprotective; noctropic; cancer;
KW antiparkinsonian; immunostimulant; cytoskeletal; immunosuppressive;
KW antidiabetic; antiallergic; gene therapy; wound healing; tissue repair;
KW burn; central nervous system disorder; Alzheimer's disease;
KW

KW Parkinson's disease; Huntington's disease; immune disorder;
 KM autoimmune disorder; multiple sclerosis; diabetes; allergy.
 XX
 OS Homo sapiens.
 XX
 PN WO200218424-A2.
 XX
 PD 07-MAR-2002.
 XX
 PF 31-AUG-2001; 2001WO-US27093.
 XX
 PR 01-SEP-2000; 2000US-0654935.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Asundi V, Zhou P, Xue AJ, Ren F, Zhang J, Wang J;
 PI Zhao QA, Wang D, Liu C, Dermanac RT, Wehrman T;
 XX
 DR WPI; 2002-583321/62.
 DR N-PDB; ABQ93334.
 XX
 PT New polynucleotide and polypeptides, useful for treatment and diagnosis
 PT of Alzheimer's, Parkinson's, Huntington's, amyotrophic lateral
 PT sclerosis, immune deficiencies, cancer, autoimmune disorders, multiple
 PT sclerosis, diabetes and allergies -
 XX
 PS Claim 20; SEQ ID NO 292; 284pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated polynucleotide (I) comprising one of
 CC 245 sequences (ABQ93288-ABQ93532). Treating a condition comprising
 CC administering to a mammalian subject a composition comprising the protein
 CC (II) encoded by (I) (ABP62809-ABP63053) or an antibody (III) to (II).
 CC (I), (II) and (III) are useful for diagnostic evaluation of disorders.
 CC (I) is useful for gene therapy of diseases and (II) can be used for
 CC therapeutic treatment. Diseases that may be treated include wound healing
 CC and tissue repair, burns, central nervous system disorders (e.g.
 CC Alzheimer's, Parkinson's, Huntington's and amyotrophic lateral
 CC sclerosis), immune deficiencies, cancer, autoimmune disorders, multiple
 CC sclerosis, diabetes and allergies.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 221 AA;
 XX

Query Match 100.0%; Score 45; DB 23; Length 221;
 Best Local Similarity 100.0%; Pred. No. 3.6; Mismatches 0; Indels 0; Gaps 0;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAA 11
 Db 156 AAAAAAAAAA 166

RESULT 15
 AAG00758
 ID AAG00758 standard; Protein: 225 AA.
 AC AAG00758;
 XX
 DT 06-OCT-2000 (first entry)
 XX
 DE Human secreted protein, SEQ ID NO: 4839.
 XX
 KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping.
 XX
 OS Homo sapiens.
 XX
 PN EPI033401-A2.
 XX
 EN 06-SEP-2000.
 XX
 PD 06-SEP-2000.
 XX

PF 21-FEB-2000; 2000EP-0200610.
 XX
 PR 26-FEB-1999; 99US-0122487.
 XX
 PA (GEST) GENSET.
 XX
 PI Dumas Milne Edwards J, Duclert A, Giordano J;
 XX
 DR WPI; 2000-500381/45.
 DR N-PDB; AAC00764.
 XX
 PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX
 PS Claim 13; SEQ ID 4839; 71pp + CD-ROM; English.
 XX
 CC The present sequence is a polypeptide encoded by one of a large number
 CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
 CC were prepared from total human RNAs or polyA+ RNAs derived from 30
 CC different tissues. EST sequences usually correspond mainly to the 3'
 CC untranslated region (UTR) of the mRNA because they are often obtained
 CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
 CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
 CC those cases where longer cDNA sequences have been obtained, the full 5'
 CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
 CC ends and can therefore be used to obtain full length cDNAs and genomic
 CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
 CC chromosome mapping procedures. They are used to obtain upstream
 CC regulatory sequences and to design expression and secretion vectors.
 XX
 SQ Sequence 225 AA;
 XX

Query Match 100.0%; Score 45; DB 21; Length 225;
 Best Local Similarity 100.0%; Pred. No. 3.7; Mismatches 0; Indels 0; Gaps 0;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAA 11
 Db 162 AAAAAAAAAA 172

Search completed: February 19, 2004, 10:03:18
 Job time : 42 secs

No data good

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 19, 2004, 10:04:34 ; Search time 34 Seconds
(without alignments)
67.741 Million cell updates/sec

Title: US-09-551-336B-1

Sequence: 1 AAAAAAAAAAAX 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 801455 seqs, 209382283 residues

Total number of hits satisfying chosen parameters: 801455

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match: 0%
Listing first 45 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/prodata/1/pubppa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/prodata/1/pubppa/US07_NEW_PUB.pep:*
- 3: /cgn2_6/prodata/1/pubppa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/prodata/1/pubppa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/prodata/1/pubppa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/prodata/1/pubppa/US08_PUBCOMB.pep:*
- 7: /cgn2_6/prodata/1/pubppa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/prodata/1/pubppa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/prodata/1/pubppa/US09_PUBCOMB.pep:*
- 10: /cgn2_6/prodata/1/pubppa/US09_PUBCOMB.pep:*
- 11: /cgn2_6/prodata/1/pubppa/US09_NEW_PUB.pep:*
- 12: /cgn2_6/prodata/1/pubppa/US10_PUBCOMB.pep:*
- 13: /cgn2_6/prodata/1/pubppa/US10_PUBCOMB.pep:*
- 14: /cgn2_6/prodata/1/pubppa/US10_PUBCOMB.pep:*
- 15: /cgn2_6/prodata/1/pubppa/US10_PUBCOMB.pep:*
- 16: /cgn2_6/prodata/1/pubppa/US10_NEW_PUB.pep:*
- 17: /cgn2_6/prodata/1/pubppa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/prodata/1/pubppa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45	100.0	20	US-10-192-832-55	Sequence 55, Appl
2	45	100.0	30	US-10-192-832-68	Sequence 68, Appl
3	45	100.0	30	US-10-192-832-69	Sequence 69, Appl
4	45	100.0	79	US-10-177-725-82	Sequence 82, Appl
5	45	100.0	79	US-10-177-725-85	Sequence 85, Appl
6	45	100.0	97	US-10-106-698-6694	Sequence 6694, Ap
7	45	100.0	104	US-10-177-725-81	Sequence 81, Appl
8	45	100.0	104	US-10-177-725-83	Sequence 83, Appl
9	45	100.0	104	US-10-177-725-86	Sequence 86, Appl
10	45	100.0	104	US-10-177-725-87	Sequence 87, Appl
11	45	100.0	104	US-10-177-725-88	Sequence 88, Appl
12	45	100.0	104	US-10-177-725-89	Sequence 89, Appl
13	45	100.0	104	US-10-177-725-90	Sequence 90, Appl
14	45	100.0	104	US-10-177-725-91	Sequence 91, Appl
15	45	100.0	104	US-10-177-725-92	Sequence 92, Appl
16	45	100.0	104	US-10-177-725-93	Sequence 93, Appl
17	45	100.0	104	US-10-177-725-94	Sequence 94, Appl
18	45	100.0	104	US-10-177-725-95	Sequence 95, Appl

16	45	100.0	104	12	US-10-177-725-81	Sequence 81, Appl
17	45	100.0	104	12	US-10-177-725-82	Sequence 82, Appl
18	45	100.0	104	12	US-10-177-725-83	Sequence 83, Appl
19	45	100.0	104	12	US-10-177-725-84	Sequence 84, Appl
20	45	100.0	104	12	US-10-177-725-85	Sequence 85, Appl
21	45	100.0	104	12	US-10-177-725-86	Sequence 86, Appl
22	45	100.0	104	12	US-10-177-725-87	Sequence 87, Appl
23	45	100.0	104	12	US-10-177-725-88	Sequence 88, Appl
24	45	100.0	104	12	US-10-177-725-89	Sequence 89, Appl
25	45	100.0	112	12	US-10-177-725-90	Sequence 90, Appl
26	45	100.0	112	12	US-10-177-725-91	Sequence 91, Appl
27	45	100.0	201	12	US-10-264-049-2597	Sequence 2597, Ap
28	45	100.0	220	9	US-09-923-304-2	Sequence 2, Appl
29	45	100.0	265	9	US-09-925-302-559	Sequence 559, Ap
30	45	100.0	617	12	US-10-104-047-2915	Sequence 2915, Ap
31	45	100.0	663	12	US-10-108-260A-2477	Sequence 2477, Ap
32	45	100.0	730	11	US-09-961-403-8	Sequence 8, Appl
33	45	100.0	731	12	US-09-964-662-1	Sequence 1, Appl
34	42	93.3	20	12	US-10-192-832-57	Sequence 2, Appl
35	42	93.3	30	12	US-10-192-832-70	Sequence 70, Appl
36	42	93.3	30	12	US-10-192-832-71	Sequence 71, Appl
37	42	93.3	102	12	US-10-121-016-9	Sequence 9, Appl
38	42	93.3	102	12	US-10-121-016-15	Sequence 15, Appl
39	42	93.3	102	12	US-10-121-016-70	Sequence 70, Appl
40	42	93.3	157	10	US-09-764-864-1325	Sequence 1325, Ap
41	42	93.3	229	12	US-10-264-049-2380	Sequence 2380, Ap
42	42	93.3	276	10	US-09-902-941-1902	Sequence 1902, Ap
43	42	93.3	276	10	US-09-849-626-1902	Sequence 1902, Ap
44	42	93.3	276	12	US-10-113-872-1902	Sequence 1902, Ap
45	42	93.3	276	15	US-10-017-754-1902	Sequence 1902, Ap

ALIGNMENTS

RESULT 1
US-10-192-832-55
; Sequence 55, Application US/10192832
; Publication No. US20030176335A1
; GENERAL INFORMATION:
; APPLICANT: ZHANG, SHUGUANG
; APPLICANT: VAUTHEY, SYLVAIN
; TITLE OF INVENTION: SUBFACTANT PEPTIDE NANOSTRUCTURES, AND USES THEREOF
; FILE REFERENCE: MTV-043.01
; CURRENT APPLICATION NUMBER: US/10/192,832
; PRIOR FILING DATE: 2002-07-10
; PRIOR APPLICATION NUMBER: 60/304,256
; PRIOR FILING DATE: 2001-07-10
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 55
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: peptide
US-10-192-832-55

Query Match 100.0%; Score 45; DB 12; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.32;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAX 11
Db 1 AAAAAAAAAAAX 11

RESULT 2
US-10-192-832-68
; Sequence 68, Application US/10192832
; Publication No. US20030176335A1
; GENERAL INFORMATION:

```

RESULT 4
US-10-177-725-22
; Sequence 22, Application US/10177725
; Publication No. US20030143562A1
; GENERAL INFORMATION:
; APPLICANT: Anderson, David
; APPLICANT: Bogenderger, Jakob M.
; APPLICANT: Peele, Beau R.
; TITLE OF INVENTION: STRUCTURALLY BIASED RANDOM PEPTIDE LIBRARIES BASED ON DIFFERENT S
; FILE REFERENCE: A-66900-4/RMS/AMS
; CURRENT APPLICATION NUMBER: US/10/177,725
; CURRENT FILING DATE: 2002-06-20

```

```

; RESULT 6
; US-10-106-698-6694
; Sequence 6694, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27

```

PRIOR APPLICATION NUMBER: PCT/US00/26524
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: US 60/157,137
PRIOR FILING DATE: 1999-09-29
PRIOR APPLICATION NUMBER: US 60/163,280
PRIOR FILING DATE: 1999-11-03
NUMBER OF SEQ ID NOS: 8564
SOFTWARE: Patentin Ver. 3.0
SEQ ID NO 6694
LENGTH: 97
TYPE: PRT
ORGANISM: Homo sapiens
US-10-106-698-6694

Query Match 100.0%; Score 45; DB 15; Length 97;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAA 11
Db 77 AAAAAAAAAA 87

RESULT 7
US-10-177-725-31
Sequence 31, Application US/10177725
Publication No. US20030143562A1
GENERAL INFORMATION:
APPLICANT: Anderson, David
APPLICANT: Bogenberger, Jakob M.
APPLICANT: Peele, Beau R.
TITLE OF INVENTION: STRUCTURALLY BIASED RANDOM PEPTIDE LIBRARIES BASED ON DIFFERENT
FILE REFERENCE: A-66900-4/RMS/AMS
CURRENT APPLICATION NUMBER: US/10/177,725
CURRENT FILING DATE: 2002-06-20
PRIOR APPLICATION NUMBER: US 09/415,765
PRIOR FILING DATE: 1999-10-08
PRIOR APPLICATION NUMBER: US 09/169,015
PRIOR FILING DATE: 1998-10-08
NUMBER OF SEQ ID NOS: 173
SOFTWARE: Patentin version 3.1
SEQ ID NO 31
LENGTH: 104
TYPE: PRT
ORGANISM: Artificial sequence
FEATURE:
OTHER INFORMATION: synthetic
US-10-177-725-31

Query Match 100.0%; Score 45; DB 12; Length 104;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAA 11
Db 93 AAAAAAAAAA 103

RESULT 8
US-10-177-725-32
Sequence 32, Application US/10177725
Publication No. US20030143562A1
GENERAL INFORMATION:
APPLICANT: Anderson, David
APPLICANT: Bogenberger, Jakob M.
APPLICANT: Peele, Beau R.
TITLE OF INVENTION: STRUCTURALLY BIASED RANDOM PEPTIDE LIBRARIES BASED ON DIFFERENT
FILE REFERENCE: A-66900-4/RMS/AMS
CURRENT APPLICATION NUMBER: US/10/177,725
CURRENT FILING DATE: 2002-06-20
PRIOR APPLICATION NUMBER: US 09/415,765
PRIOR FILING DATE: 1999-10-08
PRIOR APPLICATION NUMBER: US 09/169,015

PRIOR FILING DATE: 1998-10-08
NUMBER OF SEQ ID NOS: 173
SOFTWARE: Patentin version 3.1
SEQ ID NO 32
LENGTH: 104
TYPE: PRT
ORGANISM: Artificial sequence
FEATURE:
OTHER INFORMATION: synthetic
US-10-177-725-32

Query Match 100.0%; Score 45; DB 12; Length 104;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAA 11
Db 93 AAAAAAAAAA 103

RESULT 9
US-10-177-725-33
Sequence 33, Application US/10177725
Publication No. US20030143562A1
GENERAL INFORMATION:
APPLICANT: Anderson, David
APPLICANT: Bogenberger, Jakob M.
APPLICANT: Peele, Beau R.
TITLE OF INVENTION: STRUCTURALLY BIASED RANDOM PEPTIDE LIBRARIES BASED ON DIFFERENT
FILE REFERENCE: A-66900-4/RMS/AMS
CURRENT APPLICATION NUMBER: US/10/177,725
CURRENT FILING DATE: 2002-06-20
PRIOR APPLICATION NUMBER: US 09/415,765
PRIOR FILING DATE: 1999-10-08
PRIOR APPLICATION NUMBER: US 09/169,015
PRIOR FILING DATE: 1998-10-08
NUMBER OF SEQ ID NOS: 173
SOFTWARE: Patentin version 3.1
SEQ ID NO 33
LENGTH: 104
TYPE: PRT
ORGANISM: Artificial sequence
FEATURE:
OTHER INFORMATION: synthetic
US-10-177-725-33

Query Match 100.0%; Score 45; DB 12; Length 104;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAA 11
Db 92 AAAAAAAAAA 102

RESULT 10
US-10-177-725-34
Sequence 34, Application US/10177725
Publication No. US20030143562A1
GENERAL INFORMATION:
APPLICANT: Anderson, David
APPLICANT: Bogenberger, Jakob M.
APPLICANT: Peele, Beau R.
TITLE OF INVENTION: STRUCTURALLY BIASED RANDOM PEPTIDE LIBRARIES BASED ON DIFFERENT
FILE REFERENCE: A-66900-4/RMS/AMS
CURRENT APPLICATION NUMBER: US/10/177,725
CURRENT FILING DATE: 2002-06-20
PRIOR APPLICATION NUMBER: US 09/415,765
PRIOR FILING DATE: 1999-10-08
PRIOR APPLICATION NUMBER: US 09/169,015
PRIOR FILING DATE: 1998-10-08
NUMBER OF SEQ ID NOS: 173
SOFTWARE: Patentin version 3.1

```
; SEQ ID NO 34
; LENGTH: 104
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: synthetic
US-10-177-725-34
```

```
Query Match          100.0%; Score 45; DB 12; Length 104;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAAAK 11
Db 82 AAAAAAAAAAK 92
```

```
RESULT 11
US-10-177-725-35
; Sequence 35, Application US/10177725
; Publication No. US20030143562A1
; GENERAL INFORMATION:
; APPLICANT: Anderson, David
; APPLICANT: Bogenberger, Jakob M.
; APPLICANT: Peele, Beau R.
; TITLE OF INVENTION: STRUCTURALLY BIASED RANDOM PEPTIDE LIBRARIES BASED ON DIFFERENT S
; FILE REFERENCE: A-66900-4/RMS/AMS
; CURRENT APPLICATION NUMBER: US/10/177,725
; CURRENT FILING DATE: 2002-06-20
; PRIOR APPLICATION NUMBER: US 09/415,765
; PRIOR FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: US 09/169,015
; PRIOR FILING DATE: 1998-10-08
; NUMBER OF SEQ ID NOS: 173
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 35
; LENGTH: 104
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: synthetic
US-10-177-725-35
```

```
Query Match          100.0%; Score 45; DB 12; Length 104;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAAAK 11
Db 78 AAAAAAAAAAK 88
```

```
RESULT 12
US-10-177-725-36
; Sequence 36, Application US/10177725
; Publication No. US20030143562A1
; GENERAL INFORMATION:
; APPLICANT: Anderson, David
; APPLICANT: Bogenberger, Jakob M.
; APPLICANT: Peele, Beau R.
; TITLE OF INVENTION: STRUCTURALLY BIASED RANDOM PEPTIDE LIBRARIES BASED ON DIFFERENT S
; FILE REFERENCE: A-66900-4/RMS/AMS
; CURRENT APPLICATION NUMBER: US/10/177,725
; CURRENT FILING DATE: 2002-06-20
; PRIOR APPLICATION NUMBER: US 09/415,765
; PRIOR FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: US 09/169,015
; PRIOR FILING DATE: 1998-10-08
; NUMBER OF SEQ ID NOS: 173
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 36
; LENGTH: 104
; TYPE: PRT
```

```
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: synthetic
US-10-177-725-36
```

```
Query Match          100.0%; Score 45; DB 12; Length 104;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAAAK 11
Db 76 AAAAAAAAAAK 86
```

```
RESULT 13
US-10-177-725-37
; Sequence 37, Application US/10177725
; Publication No. US20030143562A1
; GENERAL INFORMATION:
; APPLICANT: Anderson, David
; APPLICANT: Bogenberger, Jakob M.
; APPLICANT: Peele, Beau R.
; TITLE OF INVENTION: STRUCTURALLY BIASED RANDOM PEPTIDE LIBRARIES BASED ON DIFFERENT S
; FILE REFERENCE: A-66900-4/RMS/AMS
; CURRENT APPLICATION NUMBER: US/10/177,725
; CURRENT FILING DATE: 2002-06-20
; PRIOR APPLICATION NUMBER: US 09/415,765
; PRIOR FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: US 09/169,015
; PRIOR FILING DATE: 1998-10-08
; NUMBER OF SEQ ID NOS: 173
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 37
; LENGTH: 104
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: synthetic
US-10-177-725-37
```

```
Query Match          100.0%; Score 45; DB 12; Length 104;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAAAK 11
Db 92 AAAAAAAAAAK 102
```

```
RESULT 14
US-10-177-725-38
; Sequence 38, Application US/10177725
; Publication No. US20030143562A1
; GENERAL INFORMATION:
; APPLICANT: Anderson, David
; APPLICANT: Bogenberger, Jakob M.
; APPLICANT: Peele, Beau R.
; TITLE OF INVENTION: STRUCTURALLY BIASED RANDOM PEPTIDE LIBRARIES BASED ON DIFFERENT S
; FILE REFERENCE: A-66900-4/RMS/AMS
; CURRENT APPLICATION NUMBER: US/10/177,725
; CURRENT FILING DATE: 2002-06-20
; PRIOR APPLICATION NUMBER: US 09/415,765
; PRIOR FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: US 09/169,015
; PRIOR FILING DATE: 1998-10-08
; NUMBER OF SEQ ID NOS: 173
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 38
; LENGTH: 104
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: synthetic
```

US-10-177-725-38

Query Match

100.0%; Score 45; DB 12; Length 104;

Best Local Similarity 100.0%; Pred. No. 1.5;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

1 AAAAAAAAAAK 11

Db

92 AAAAAAAAAAK 102

RESULT 15

US-10-177-725-42

; Sequence 42, Application US/10177725

; Publication No. US20030143562A1

; GENERAL INFORMATION:

; APPLICANT: Anderson, David

; APPLICANT: Bogenberger, Jakob M.

; APPLICANT: Peele, Beau R.

; TITLE OF INVENTION: STRUCTURALLY BIASED RANDOM PEPTIDE LIBRARIES BASED ON DIFFERENT S

; FILE REFERENCE: A-66900-4/RMS/AMS

; CURRENT APPLICATION NUMBER: US/10/177,725

; CURRENT FILING DATE: 2002-06-20

; PRIOR APPLICATION NUMBER: US 09/415,765

; PRIOR FILING DATE: 1999-10-08

; PRIOR APPLICATION NUMBER: US 09/169,015

; PRIOR FILING DATE: 1998-10-08

; NUMBER OF SEQ ID NOS: 173

; SOFTWARE: Patentin version 3.1

; SEQ ID NO 42

; LENGTH: 104

; TYPE: PRT

; ORGANISM: Artificial sequence

; FEATURE:

; OTHER INFORMATION: synthetic

US-10-177-725-42

Query Match

100.0%; Score 45; DB 12; Length 104;

Best Local Similarity 100.0%; Pred. No. 1.5;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

1 AAAAAAAAAAK 11

Db

92 AAAAAAAAAAK 102

Search completed: February 19, 2004, 10:10:03
Job time : 35 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: February 19, 2004, 10:02:28 / Search time 22 Seconds
(without alignments)
21.155 Million cell updates/sec

Title: US-09-551-336B-1
Perfect score: 45
Sequence: 1 AAAAAAAAAAK 11

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 328717 seqs, 42310858 residues
Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database:
- 1: /cgn2_6/prodata/1/iaa/5A_COMB.pep.*
 - 2: /cgn2_6/prodata/1/iaa/5B_COMB.pep.*
 - 3: /cgn2_6/prodata/1/iaa/5A_COMB.pep.*
 - 4: /cgn2_6/prodata/1/iaa/5B_COMB.pep.*
 - 5: /cgn2_6/prodata/1/iaa/5A_COMB.pep.*
 - 6: /cgn2_6/prodata/1/iaa/5B_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45	100.0	731	2	US-08-911-364-1
2	45	100.0	731	4	US-09-340-736E-1
3	45	100.0	733	3	US-08-464-700-2
4	45	100.0	792	2	US-08-678-039A-40
5	42	93.3	151	3	US-09-045-764A-4
6	41	91.1	39	4	US-09-117-121-28
7	41	91.1	54	4	US-09-117-121-30
8	41	91.1	238	3	US-09-234-332-10
9	41	91.1	238	4	US-09-702-705-1813
10	41	91.1	238	4	US-09-716-457-1813
11	41	91.1	340	3	US-09-360-779-2
12	41	91.1	340	4	US-09-435-335-2
13	41	91.1	11	3	US-09-208-966-53
14	40	88.9	12	1	US-07-392-288-7
15	40	88.9	12	1	US-07-989-764-7
16	40	88.9	12	3	US-09-058-459-37
17	40	88.9	12	3	US-09-127-926-37
18	40	88.9	13	4	US-08-788-822A-15
19	40	88.9	14	3	US-09-058-562-31
20	40	88.9	18	1	US-08-240-712-33
21	40	88.9	22	1	US-08-443-890-33
22	40	88.9	22	1	US-07-992-288-6
23	40	88.9	22	1	US-07-989-764-6
24	40	88.9	37	1	US-08-209-747-49
25	40	88.9	37	1	US-08-458-298-49
26	40	88.9	47	1	US-08-425-069-26
27	40	88.9	47	1	US-08-209-747-37

28	40	88.9	47	1	US-08-458-298-37	Sequence 37, Appl
29	40	88.9	47	2	US-08-317-844B-26	Sequence 26, Appl
30	40	88.9	47	3	US-08-556-978B-18	Sequence 18, Appl
31	40	88.9	177	3	US-09-058-562-27	Sequence 27, Appl
32	40	88.9	182	1	US-08-240-712-29	Sequence 29, Appl
33	40	88.9	182	1	US-08-443-890-29	Sequence 29, Appl
34	40	88.9	223	4	US-09-009-816-4	Sequence 4, Appl
35	40	88.9	238	4	US-09-140-749-51	Sequence 51, Appl
36	40	88.9	367	4	US-09-009-816-2	Sequence 2, Appl
37	40	88.9	407	2	US-08-765-875-2	Sequence 2, Appl
38	40	88.9	407	2	US-08-765-875-6	Sequence 6, Appl
39	40	88.9	407	3	US-08-795-671-2	Sequence 2, Appl
40	40	88.9	407	3	US-08-795-671-6	Sequence 6, Appl
41	40	88.9	407	4	US-09-454-540-2	Sequence 4, Appl
42	40	88.9	407	4	US-09-454-540-6	Sequence 6, Appl
43	40	88.9	442	3	US-09-347-833-11	Sequence 11, Appl
44	40	88.9	449	2	US-08-927-394-2	Sequence 2, Appl
45	40	88.9	453	6	5206152-7	Patent No. 5206152

ALIGNMENTS

RESULT 1
US-08-911-364-1
Sequence 1 Application US/08911364
Patent No 5265106

GENERAL INFORMATION:

APPLICANT: ROTHSTEIN, Aser
APPLICANT: KEELY, Fred W.
TITLE OF INVENTION: SELF-ALIGNING PEPTIDES MODELED ON HUMAN
TITLE OF INVENTION: ELASTIN AND OTHER FIBROUS PROTEINS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY & LARDNER
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/911,364
FILING DATE: 07-AUG-1997
CLASSIFICATION: 530
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 60/023,552
FILING DATE: 07-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Bent, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 041082/0104
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 731 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-911-364-1

8/7/97 Good

Query Match 100.0%, Score 45, DB 2, Length 731;
Best Local Similarity 100.0%, Pred. No. 1.8, Indels 0, Gaps 0;
Matches 11, Conservative 0, Mismatches 0, Indels 0, Gaps 0;

QY 1 AAAAAAAAAA 11
DB 276 AAAAAAAAAA 286

RESULT 2
US-09-340-736E-1
Sequence 2, Application US/09340736E
Patent No. 6489446
GENERAL INFORMATION:
APPLICANT: ROTHSTEIN, ASER
APPLICANT: KEELEY, FRED
APPLICANT: ROTHSTEIN, STEVEN
TITLE OF INVENTION: SELF-ALIGNING PEPTIDES MODELLED ON HUMAN ELASTIN
TITLE OF INVENTION: AND OTHER FIBROUS PROTEINS
FILE REFERENCE: 041082/0110
CURRENT APPLICATION NUMBER: US/09/340,736E
CURRENT FILING DATE: 1999-06-29
PRIOR APPLICATION NUMBER: 08/911,364
PRIOR FILING DATE: 1997-08-07
PRIOR APPLICATION NUMBER: 60/023,552
PRIOR FILING DATE: 1996-08-07
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 731
TYPE: PRT
ORGANISM: Homo sapiens
US-09-340-736E-1

Query Match 100.0%; Score 45; DB 4; Length 731;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAA 11
DB 276 AAAAAAAAAA 286

RESULT 3
US-08-678-039A-2
Sequence 2, Application US/08464700
Patent No. 6232458
GENERAL INFORMATION:
APPLICANT: WEISS, ANTHONY S
APPLICANT: MARTIN, STEPHEN L
TITLE OF INVENTION: SYNTHETIC POLYNUCLEOTIDES
NUMBER OF SEQUENCES: 54
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Cntr, PO Box 457
CITY: Spring House
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/464,700
FILING DATE: 7-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PL6520
FILING DATE: 22-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PL9661
FILING DATE: 28-JUN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU93/00655
FILING DATE: 16-DEC-1993

ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: GHCUUSA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
INFORMATION FOR SEQ ID NO: 2
SEQUENCE CHARACTERISTICS:
LENGTH: 733 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-464-700-2

Query Match 100.0%; Score 45; DB 3; Length 733;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAA 11
DB 278 AAAAAAAAAA 288

RESULT 4
US-08-678-039A-40
Sequence 40, Application US/08678039A
Patent No. 5858662
GENERAL INFORMATION:
APPLICANT: Keating, Mark T.
APPLICANT: Morris, Colleen A.
TITLE OF INVENTION: Diagnosis of Williams Syndrome and
TITLE OF INVENTION: Williams Syndrome Cognitive Profile by Analysis of the
TITLE OF INVENTION: Presence or Absence of a LIM-Kinase Gene
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Rothwell, Figg, Ernst & Kurtz, P.C.
STREET: 555 Thirteenth Street, N.W., Suite 701 East
STREET: Tower
CITY: Washington
STATE: DC
COUNTRY: U.S.A.
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/678,039A
FILING DATE: 10-JUL-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Saxe, Stephen A.
REGISTRATION NUMBER: 38,609
REFERENCE/DOCKET NUMBER: 2323-120A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-624-1589
TELEFAX: 202-783-6031
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 792 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-678-039A-40

Query Match 100.0%; Score 45; DB 2; Length 792;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAA 11

Db 302 AAAAAAAAAAK 312

RESULT 5

US-09-045-764A-4

; Sequence 4, Application US/09045764A

; Patent No. 6127178

; GENERAL INFORMATION:

; APPLICANT: Israel, Mark A.

; APPLICANT: Florio, Monica

; TITLE OF INVENTION: Apoptotic Peptides

; NUMBER OF SEQUENCES: 12

; CORRESPONDENCE ADDRESSES:

; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP

; STREET: 75 DENISE DRIVE

; CITY: HILLSBOROUGH

; STATE: CALIFORNIA

; COUNTRY: USA

; ZIP: 94010

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/045,764A

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: OSMAN, RICHARD A

; REGISTRATION NUMBER: 36,627

; REFERENCE/DOCKET NUMBER: UCSF98-045

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (650) 343-4341

; TELEFAX: (650) 343-4342

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 161 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

US-09-045-764A-4

Query Match 93.3%; Score 42; DB 3; Length 161;

Best Local Similarity 90.9%; Pred. No. 1.2;

Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAK 11

Db 39 AAAAAAAAAAK 49

RESULT 6

US-09-117-121-28

; Sequence 28, Application US/09117121

; Patent No. 6307020

; GENERAL INFORMATION:

; APPLICANT: Hew, Choy

; APPLICANT: Gong, Zhiyuan

; TITLE OF INVENTION: Intracellular Antifreeze Polypeptides

; TITLE OF INVENTION: and Nucleic Acids

; NUMBER OF SEQUENCES: 46

; CORRESPONDENCE ADDRESSES:

; ADDRESSEE: Townsend and Townsend and Crew LLP

; STREET: Two Embarcadero Center, Eighth Floor

; CITY: San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94111-3834

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/117,121

; FILING DATE: 20-NOV-1998

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: WO PCT/CA97/00062

; FILING DATE: 30-JAN-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Weber, Kenneth A.

; REGISTRATION NUMBER: 31,677

; REFERENCE/DOCKET NUMBER: 016252-001610US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 576-0200

; TELEFAX: (415) 576-0300

; INFORMATION FOR SEQ ID NO: 28:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 39 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-09-117-121-28

Query Match 91.1%; Score 41; DB 4; Length 39;

Best Local Similarity 90.9%; Pred. No. 0.45;

Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAK 11

Db 6 AAAAAATAAAK 16

RESULT 7

US-09-117-121-30

; Sequence 30, Application US/09117121

; Patent No. 6307020

; GENERAL INFORMATION:

; APPLICANT: Hew, Choy

; APPLICANT: Gong, Zhiyuan

; TITLE OF INVENTION: Intracellular Antifreeze Polypeptides

; TITLE OF INVENTION: and Nucleic Acids

; NUMBER OF SEQUENCES: 46

; CORRESPONDENCE ADDRESSES:

; ADDRESSEE: Townsend and Townsend and Crew LLP

; STREET: Two Embarcadero Center, Eighth Floor

; CITY: San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94111-3834

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/117,121

; FILING DATE: 20-NOV-1998

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: WO PCT/CA97/00062

; FILING DATE: 30-JAN-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Weber, Kenneth A.

; REGISTRATION NUMBER: 31,677

; REFERENCE/DOCKET NUMBER: 016252-001610US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 576-0200

; TELEFAX: (415) 576-0300

; INFORMATION FOR SEQ ID NO: 30:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 54 amino acids

; TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-117-121-30

Query Match 91.1%; Score 41; DB 4; Length 54;
Best Local Similarity 90.9%; Pred. No. 0.62;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAA 11
Db 6 AAAAAAAAAA 16

RESULT 8

US-09-234-332-10
Sequence 10, Application US/09234332A
Patent No. 6087168
GENERAL INFORMATION:
APPLICANT: Cedars-Sinai Medical Center
APPLICANT: Michel F. Levesque, M.D.
APPLICANT: Thomas Neuman, Ph.D.
TITLE OF INVENTION: CONVERSION OF NON-NEURONAL CELLS INTO
FILE REFERENCE: P07 41494
CURRENT APPLICATION NUMBER: US/09/234,332A
CURRENT FILING DATE: 1999-01-20
NUMBER OF SEQ ID NOS: 16
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 10
LENGTH: 238
TYPE: PRT
ORGANISM: Human
FEATURE:
NAME/KEY: PEPTIDE
LOCATION: (0)...(0)
OTHER INFORMATION: Achete scute homologous protein (ASH1); Genbank
US-09-234-332-10

Query Match 91.1%; Score 41; DB 3; Length 238;
Best Local Similarity 90.9%; Pred. No. 2.7;
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAA 11
Db 38 AAAAAAAAAA 48

RESULT 9

US-09-702-705-1813
Sequence 1813, Application US/09702705
Patent No. 6504010
GENERAL INFORMATION:
APPLICANT: Wang, Tongtong
APPLICANT: Bangur, Chaitanya S.
APPLICANT: Lodes, Michael A.
APPLICANT: Fanger, Gary
APPLICANT: Vedvick, Tom
APPLICANT: Carter, Darrick
APPLICANT: Retter, Marc
APPLICANT: Mannion, Jane
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
FILE REFERENCE: 210121.478C14
CURRENT APPLICATION NUMBER: US/09/702,705
CURRENT FILING DATE: 2000-10-30
NUMBER OF SEQ ID NOS: 1833
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1813
LENGTH: 238
TYPE: PRT
ORGANISM: Homo sapiens

US-09-702-705-1813

Query Match 91.1%; Score 41; DB 4; Length 238;
Best Local Similarity 90.9%; Pred. No. 2.7;
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAA 11
Db 38 AAAAAAAAAA 48

RESULT 10

US-09-736-457-1813
Sequence 1813, Application US/09736457
Patent No. 6509448
GENERAL INFORMATION:
APPLICANT: Wang, Tongtong
APPLICANT: Bangur, Chaitanya S.
APPLICANT: Lodes, Michael A.
APPLICANT: Fanger, Gary
APPLICANT: Vedvick, Tom
APPLICANT: Carter, Darrick
APPLICANT: Retter, Marc
APPLICANT: Mannion, Jane
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
FILE REFERENCE: 210121.478C15
CURRENT APPLICATION NUMBER: US/09/736,457
CURRENT FILING DATE: 2000-12-13
NUMBER OF SEQ ID NOS: 1864
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1813
LENGTH: 238
TYPE: PRT
ORGANISM: Homo sapiens
US-09-736-457-1813

Query Match 91.1%; Score 41; DB 4; Length 238;
Best Local Similarity 90.9%; Pred. No. 2.7;
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAA 11
Db 38 AAAAAAAAAA 48

RESULT 11

US-09-360-779-2
Sequence 2, Application US/09360779
Patent No. 6268216
GENERAL INFORMATION:
APPLICANT: Deneris, Evan S.
APPLICANT: Eyodoro, Dmitry V.
APPLICANT: Hendricks, Timothy J.
TITLE OF INVENTION: Reagents and Methods for the Screening of Compounds
TITLE OF INVENTION: Useful in the Treatment of Neurological Diseases
FILE REFERENCE: CASE-03828
CURRENT APPLICATION NUMBER: US/09/360,779
CURRENT FILING DATE: 1999-07-26
EARLIER APPLICATION NUMBER: 60/094,264
EARLIER FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 340
TYPE: PRT
ORGANISM: Rattus norvegicus
US-09-360-779-2

Query Match 91.1%; Score 41; DB 3; Length 340;
Best Local Similarity 90.9%; Pred. No. 3.9;

Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAA 11
| | | | | | | | | |
Db 249 AAAAAAAAAA 259

RESULT 12

US-09-435-335-2
; Sequence 2, Application US/09435335
; Patent No. 6384204
; GENERAL INFORMATION:
; APPLICANT: Denetis, Evan S.
; APPLICANT: Eyodoro, Dmitry V.
; APPLICANT: Hendricks, Timothy J.
; TITLE OF INVENTION: Reagents and Methods for the Screening of Compounds
; FILE REFERENCE: CASE-04027
; CURRENT APPLICATION NUMBER: US/09/435,335
; EARLIER FILING DATE: 1999-11-05
; EARLIER APPLICATION NUMBER: 09/360,779
; EARLIER FILING DATE: 1999-07-26
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 340
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-09-435-335-2

Query Match 91.1%; Score 41; DB 4; Length 340;
Best Local Similarity 90.9%; Pred. No. 3.9;
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAA 11
| | | | | | | | | |
Db 249 AAAAAAAAAA 259

RESULT 13

US-09-208-966-53
; Sequence 53, Application US/09208966
; Patent No. 6221355
; GENERAL INFORMATION:
; APPLICANT: Dowdy, Steven F.
; TITLE OF INVENTION: ANT-PATHOGEN SYSTEM AND METHODS OF USE THEREOF
; FILE REFERENCE: 48861/1742
; CURRENT APPLICATION NUMBER: US/09/208,966
; CURRENT FILING DATE: 1998-12-10
; EARLIER APPLICATION NUMBER: 60/082,402
; EARLIER FILING DATE: 1998-04-20
; EARLIER APPLICATION NUMBER: 60/069,012
; EARLIER FILING DATE: 1997-12-10
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 53
; LENGTH: 11
; TYPE: PRT
; ORGANISM: human
US-09-208-966-53

Query Match 88.9%; Score 40; DB 3; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAA 10
| | | | | | | | | |
Db 2 AAAAAAAAAA 11

RESULT 14
US-07-992-288-7
; Sequence 7, Application US/07992288

; Patent No. 5338831
; GENERAL INFORMATION:
; APPLICANT: Lebel, Michel
; APPLICANT: Richler, Jutta
; APPLICANT: Pokorny, Vite
; APPLICANT: Jehnicka, Jiri
; APPLICANT: Mudra, Petr
; APPLICANT: Zenisek, Karel
; APPLICANT: Stierandova, Alena
; APPLICANT: Kalousek, Jan
; APPLICANT: Bolf, Jan
; TITLE OF INVENTION: METHOD OF MAKING MULTIPLE SYNTHESIS OF
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESS: Dressler, Goldsmith, Shore & Milnamow, Ltd.
; STREET: 180 No. 5338831th Stetson, Suite 4700
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/992,288
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/645,121
; FILING DATE: 24-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Hoover, Allen J.
; REGISTRATION NUMBER: 24,103
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 616-5400
; TELEFAX: (312) 616-5460
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-07-992-288-7

Query Match 88.9%; Score 40; DB 1; Length 12;
Best Local Similarity 100.0%; Pred. No. 0.2;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAA 10
| | | | | | | | | |
Db 1 AAAAAAAAAA 10

RESULT 15
US-07-989-764-7
; Sequence 7, Application US/07989764
; Patent No. 5342585
; GENERAL INFORMATION:
; APPLICANT: Lebel, Michel
; APPLICANT: Richler, Jutta
; APPLICANT: Pokorny, Vite
; APPLICANT: Jehnicka, Jiri
; APPLICANT: Mudra, Petr
; APPLICANT: Zenisek, Karel
; APPLICANT: Stierandova, Alena
; APPLICANT: Kalousek, Jan
; APPLICANT: Bolf, Jan
; TITLE OF INVENTION: APPARATUS FOR MAKING MULTIPLE SYNTHESIS
; TITLE OF INVENTION: OF PEPTIDES ON SOLID SUPPORT
; NUMBER OF SEQUENCES: 7

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Dressler, Goldsmith, Shore & Milnamow, Ltd.
;; STREET: 180 No. 5342585th Stetson, Suite 4700
;; City: Chicago
;; STATE: IL
;; COUNTRY: USA
;; ZIP: 60601
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent In Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/07/989,764
;; FILING DATE:
;; CLASSIFICATION: 530
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/645,121
;; FILING DATE: 24-JAN-1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Hoover, Allen J.
;; REGISTRATION NUMBER: 24,103
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (312)616-5400
;; TELEFAX: (312)616-5460
;; INFORMATION FOR SEQ ID NO: 7:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 12 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
;; US-07-989-764-7

Query Match 88.9%; Score 40; DB 1; Length 12;
Best Local Similarity 100.0%; Pred. No. 0.2;
Matches 10; Conservative 0; Mismatches 0; Indels 0;
Gaps 0;

Qy 1 AAAAAAAAAA 10
Db 1 AAAAAAAAAA 10

Search completed: February 19, 2004, 10:05:37
Job time : 23 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Comphen Ltd.

OM protein - protein search, using sw model

Run on: February 19, 2004, 10:01:38 / Search time 21 Seconds

(without alignments)
50.374 Million cell updates/sec

Title: US-09-551-336B-1

Perfect score: 45

Sequence: 1 AAAAAAAAAAK 11

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

1: p1r1:*

2: p1r2:*

3: p1r3:*

4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	100.0	220	2 JC5954	ribosomal protein
2	45	100.0	435	2 A44308	Antho-RFamide prec
3	45	100.0	792	1 EAHU	elastin precursor,
4	45	100.0	1430	2 T34516	hypothetical prote
5	42	93.3	161	2 JEO306	Id4 protein - huma
6	42	93.3	161	2 G01855	helix-loop-helix p
7	42	93.3	161	2 S43260	unknown protein F9
8	42	93.3	333	2 G96780	hypothetical prote
9	42	93.3	814	2 T47641	serine-rich protei
10	42	93.3	1077	2 A44067	abdominal segment
11	42	93.3	1533	2 A46221	long chain fatty a
12	42	93.3	1607	2 T02837	fibroin - Chinese
13	42	93.3	2639	2 T31328	homoeotic protein u
14	41	91.1	40	2 S58853	achae-scute locu
15	41	91.1	97	2 S02376	probable MASH-2 pr
16	41	91.1	231	2 S28186	hypothetical prote
17	41	91.1	233	2 S11563	hypothetical prote
18	41	91.1	238	2 T11718	paired type homeob
19	41	91.1	238	2 A48279	homoeotic protein H
20	41	91.1	273	2 T51010	segmentation prote
21	41	91.1	287	2 T51011	homoeotic protein e
22	41	91.1	314	2 JC5273	transcription fact
23	41	91.1	323	2 S16318	protein F53A3.6 (l
24	41	91.1	337	2 S06956	arylsulfatase (EC
25	41	91.1	392	2 B48423	transcription fact
26	41	91.1	451	1 A40168	
27	41	91.1	451	1 D88395	
28	41	91.1	649	2 S43229	
29	41	91.1	701	1 S46458	

30	41	91.1	702	2 G01840	T-box protein 2 -
31	41	91.1	747	1 EABO	elastin precursor,
32	41	91.1	748	1 T49633	glucan 1,4-alpha-g
33	41	91.1	770	2 S59623	tropoelastin - she
34	41	91.1	796	2 UC7555	C1orf4 protein -
35	41	91.1	860	1 EAMS	elastin precursor
36	41	91.1	864	1 EART	elastin precursor
37	41	91.1	1065	2 T13230	dachshund isoform
38	41	91.1	1072	2 T13232	dachshund protein
39	41	91.1	1074	2 T13229	dachshund protein
40	41	91.1	1081	2 T13231	dachshund protein
41	41	91.1	1175	2 JH0697	potassium channel
42	41	91.1	1180	2 S69205	stripe a/b protein
43	41	91.1	1184	2 A39800	calcium-activated
44	41	91.1	1355	2 S40022	spalt protein - fr
45	41	91.1	1402	2 S42748	finger protein - f

ALIGNMENTS

RESULT 1

UC5954 ribosomal protein L14 - human
C:Species: Homo sapiens (man)
C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 02-Sep-2000
C:Accession: JC5954
R:Tanaka, M.; Tanaka, T.; Harata, M.; Suzuki, T.; Mitsui, Y.
Biochem. Biophys. Res. Commun. 243, 531-537, 1998
A>Title: Triplet repeat-containing ribosomal protein L14 gene in immortalized human endo
A:Reference number: JC5954; MUID:98153799; PMID:9480843
A:Accession: JC5954
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-220 <TAN>
A:Cross-references: DDBJ:D87735; NID:G1620021; PIDN:BA113443.1; PID:G1620022
C:Superfamily: rat ribosomal protein L14

Query Match

Best Local Similarity 100.0%; Score 45; DB 2; Length 220;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAK 11
Db 155 AAAAAAAAAAK 165

RESULT 2
A4308 Antho-RFamide precursor - sea anemone (Anthopleura elegantissima)

C:Species: Anthopleura elegantissima
C>Date: (0-Jun-1995) #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000

R:Schmuckler, C.; Darner, D.; Diekhoff, D.; Grimmelikhuijzen, C.J.

J. Biol. Chem. 267, 22534-22541, 1992

A:Title: Identification of a novel type of processing sites in the precursor for the

A:Reference number: A4308; MUID:93054550; PMID:1425603

A:Accession: A4308

A>Status: preliminary

A:Molecule type: mRNA; protein

A:Residues: 1-435 <SCH>

A:Cross-references: GB:M68269; NID:G155702; PIDN:AAA27738.1; PID:G155703

C:Keywords: neuropeptide

Qy 1 AAAAAAAAAAK 11
Db 377 AAAAAAAAAAK 387

Query Match 100.0%; Score 45; DB 2; Length 435;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 3

EARTH

elastin precursor, long splice form - human

N/Alternate names: tropoelastin

C/Species: Homo sapiens (man)

C/Date: 22-Jun-1990 #sequence, revision 26-Jul-1996 #text_change 22-Jun-1999

C/Accession: A32707; A33705; A30524; A53891

R/Indik, Z.; Yeh, H.; Ornstein-Goldstein, N.; Sheppard, P.; Anderson, N.; Rosenbloom, J.

Proc. Natl. Acad. Sci. U.S.A. 84, 5680-5684, 1987

A/Title: Alternative splicing of human elastin mRNA indicated by sequence analysis of cDNA

A/Reference number: A32707; MUID:87289668; PMID:3039501

A/Accession: A32707

A/Molecule type: mRNA

A/Residues: 1-500,507-792 <IND>

A/Cross-references: GB:M16983; GB:J02948

R/Bashir, M.M.; Indik, Z.; Yeh, H.; Ornstein-Goldstein, N.; Rosenbloom, J.C.; Abrams, W.

J. Biol. Chem. 264, 8887-8891, 1989

A/Title: Characterization of the complete human elastin gene. Delineation of unusual features

A/Reference number: A33705; MUID:89255358; PMID:2722804

A/Accession: A33705

A/Molecule type: DNA

A/Residues: 1-27 <BAS>

A/Cross-references: GB:J04821; NID:g182052; PIDN:AAA52379.1; PID:g553276

R/Fazio, M.J.; Olsen, D.R.; Kaub, E.A.; Baldwin, C.T.; Indik, Z.; Ornstein-Goldstein, N.

J. Invest. Dermatol. 91, 458-464, 1988

A/Title: Cloning of full-length elastin cDNAs from a human skin fibroblast recombinant

A/Reference number: A30524; MUID:89009960; PMID:3171221

A/Accession: A30524

A/Molecule type: mRNA

A/Residues: 1-453,483-617,651-792 <FA2>

A/Cross-references: EMBL:M56860; NID:g182061; PIDN:AAA52382.1; PID:g182062

R/Fazio, M.J.; Olsen, D.R.; Kuitavien, H.; Chu, M.L.; Davidson, J.M.; Rosenbloom, J.; U

Lab. Invest. 58, 270-277, 1988

A/Title: Isolation and characterization of human elastin cDNAs, and age-associated varia

A/Reference number: A53891; MUID:88156138; PMID:2831431

A/Accession: A53891

A/Molecule type: mRNA

A/Residues: 164-453,483-500,507-617,651-792 <FA2>

A/Cross-references: GB:M24782; NID:g182063; PIDN:AAA53190.1; PID:g182064

C/Comment: The term tropoelastin refers to a soluble precursor form of the extracellular

line oxidase activity.

A/Status: preliminary; translated from GB/EMBL/DDBJ

A/Molecule type: DNA

A/Residues: 1-1430 <FAV>

A/Cross-references: EMBL:U13646; PIDN:AAC24421.1; GSPDB:GN00021; CESP:ZK783.4

A/Experimental source: strain Bristol N2; clone ZK783

C/Species: Homo sapiens (man)

C/Date: 21-Dec-1996 #sequence, revision 06-Jun-1997 #text_change 19-May-2000

C/Accession: G01855

R/Kiesling, T.L.

submitted to the EMBL Data Library, June 1995

A/Reference number: G08632

A/Accession: G01855

A/Molecule type: translated from GB/EMBL/DDBJ

A/Residues: 1-161 <KIE>

A/Cross-references: EMBL:U28368; NID:g881545; PIDN:AAA73923.1; PID:g881546

C/Suprafamily: transcription repressor Id-2

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-161 <RIG>

A/Cross-references: GB:Y07958

C/Suprafamily: transcription repressor Id-2

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-161 <RIG>

A/Cross-references: GB:Y07958

C/Suprafamily: transcription repressor Id-2

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-161 <RIG>

A/Cross-references: GB:Y07958

C/Suprafamily: transcription repressor Id-2

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-161 <RIG>

A/Cross-references: GB:Y07958

C/Suprafamily: transcription repressor Id-2

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-161 <RIG>

A/Cross-references: GB:Y07958

S43260
 helix-loop-helix protein Id4, dominant negative - mouse
 A:Species: Mus musculus (house mouse)
 C:Date: 07-Oct-1994 #sequence_revision 07-Oct-1994 #text_change 24-Nov-1999
 C:Accession: S43260; S37199
 R:Riechmann, V.; van Cruchten, I.; Sablitzky, F.
 Nucleic Acids Res. 22, 749-755, 1994
 A:Title: The expression pattern of Id4, a novel dominant negative helix-loop-helix prote
 A:Reference number: S43260; MUID:94188125; PMID:8139914
 A:Accession: S43260
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-161 <RIB>
 A:Cross-references: EMBL:X75018; NID:g402637; PIDN:CAA52926.1; PID:g402638
 C:Superfamily: transcription repressor Id-2
 F:64-106/Region: helix-loop-helix #status predicted

Query Match 93.3%; Score 42; DB 2; Length 161;
 Best Local Similarity 90.9%; Pred. No. 15;
 Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAA 11
 |||||
 Db 39 AAAAAAAAAA 49

RESULT 8

G96780
 unknown protein F9E10.7 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: G96780
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Coml, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
 ansen, N.F.; Hughes, B.; Holtz, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Malt, R.; Marziani,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 Ker, M.; Wu, D.; Yu, G.; Frazer, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719; PMID:11130712
 A:Accession: G96780
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-333 <STO>
 A:Cross-references: GB:AE005173; NID:g6646757; PIDN:AAF21069.1; GSPDB:GN00141
 C:Genetics:
 A:Gene: F9E10.7
 A:Map position: 1

Query Match 93.3%; Score 42; DB 2; Length 333;
 Best Local Similarity 90.9%; Pred. No. 25;
 Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAA 11
 |||||
 Db 12 AAAAAAAAAA 22

RESULT 9

T47641
 hypothetical protein T15C9.20 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
 C:Accession: T47641
 R:Mewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
 submitted to the Protein Sequence Database, April 2000
 A:Reference number: Z24470
 A:Accession: T47641
 A:Status: preliminary
 A:Molecule type: DNA

A:Residues: 1-814 <MEM>
 A:Cross-references: EMBL:AL132970
 A:Experimental source: cultivar Columbia; BAC clone T15C9
 C:Genetics:
 A:Map position: 3
 A:Intron: 28/2; 51/3; 252/3; 312/3; 350/3; 370/2; 392/3; 410/1; 441/1; 469/1; 604/3; 62
 A:Note: T15C9.20

Query Match 93.3%; Score 42; DB 2; Length 814;
 Best Local Similarity 90.9%; Pred. No. 49;
 Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAA 11
 |||||
 Db 6 AAAAAAAAAA 16

RESULT 10

A44067
 serine-rich protein hairless - fruit fly (Drosophila melanogaster)
 N:Alternate names: 109K basic protein H
 C:Species: Drosophila melanogaster
 C:Date: 10-Jun-1993 #sequence_revision 26-Feb-1999 #text_change 26-Feb-1999
 C:Accession: A44067; A58929; S33412; S24639
 R:Bang, A.G.; Posakony, J.W.
 Genes Dev. 6, 1752-1769, 1992
 A:Title: The Drosophila gene Hairless encodes a novel basic protein that controls altern
 A:Reference number: A44067; MUID:92387549; PMID:1516831
 A:Accession: A44067
 A:Molecule type: DNA
 A:Residues: 19-1077 <BAN>
 A:Cross-references: GB:M95192; NID:g157621; PID:g157622
 A:Note: sequence extracted from NCBI backbone (NCBIN:112622, NCBI:112623)
 R:Preiss, A.
 submitted to the EMBL Data Library, May 1994
 A:Description: Hairless, a Drosophila gene involved in neural development, encodes a nov
 A:Reference number: A58929
 A:Accession: A58929
 A:Molecule type: mRNA
 A:Residues: 1-1077 <PRE>
 A:Cross-references: EMBL:X67239; GB:S49642; NID:9578331; PID:9578332
 R:Maier, D.; Stumm, G.; Kuhn, K.; Preiss, A.
 Mech. Dev. 38, 143-156, 1992
 A:Title: Hairless, a Drosophila gene involved in neural development, encodes a novel, se
 A:Reference number: S33412; MUID:93041287; PMID:1419850
 A:Accession: S33412
 A:Molecule type: mRNA
 A:Residues: 1-150, 'A', 152-701, 'LV', 704-890, 'R', 892-963, 'RLIP', 968-973, 975-1077 <MAI>
 A:Cross-references: EMBL:X67239
 C:Genetics:
 A:Gene: FlyBase:H; hairless
 A:Cross-references: FlyBase:FBgn0001169

Query Match 93.3%; Score 42; DB 2; Length 1077;
 Best Local Similarity 90.9%; Pred. No. 61;
 Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAA 11
 |||||
 Db 937 AAAAAAAAAA 947

RESULT 11

A46221
 abdominal segment formation protein puntillo - fruit fly (Drosophila melanogaster)
 C:Species: Drosophila melanogaster
 C:Date: 21-Sep-1993 #sequence_revision 25-Apr-1997 #text_change 17-Oct-1997
 C:Accession: A46221; S22026
 R:Barker, D.D.; Wang, C.; Moore, J.; Dickinson, L.K.; Lehmann, R.
 Genes Dev. 6, 2312-2326, 1992
 A:Title: Puntillo is essential for function but not for distribution of the Drosophila at
 A:Reference number: A46221; MUID:93093466; PMID:1459455

A/Accession: A46221
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-1533 <BAR>
 A/Cross-references: GB:U07943; NID:g158190; PID:g158191
 A/Note: sequence extracted from NCBI backbone (NCBIN:120203, NCBIPI:120204)
 R/Macdonald, P.M.
 Submitted to the EMBL Data Library, October 1991
 A/Reference number: S22026
 A/Accession: S22026
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-361, 'A', 363-1102, 'R', 1104-1405, 'NN', 1408-1495, 'V', 1497-1518, 'S', 1520-1533
 A/Cross-references: EMBL:X62589; NID:g8393; PID:g8394
 C/Genetics:
 A/Gene: FlyBase:pum
 A/Cross-references: FlyBase:FBgn0003165

Query Match 93.3%; Score 42; DB 2; Length 1533;
 Best Local Similarity 90.9%; Pred. No. 79;
 Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAK 11
 |||||
 Db 937 AAAAAAAAAAR 947

RESULT 12
 T02837
 long chain fatty acyl CoA synthetase LCPACAS5 [imported] - Leishmania major (strain Friedl
 C/Species: Leishmania major
 C/Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 19-May-2000
 C/Accession: A81461; T02837
 R/Mayer, P.J.; Audleman, L.; deVos, T.; Hixson, G.; Kiser, P.; Lemley, C.; Magness, C.;
 Proc. Natl. Acad. Sci. U.S.A. 96, 2902-2906, 1999
 A/Title: Leishmania major Friedlin chromosome 1 has an unusual distribution of protein-c
 A/Reference number: A81455; MUID:99178987; PMID:10077609
 A/Accession: A81461
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-1607 <EYL>
 A/Cross-references: GB:AE001274; NID:g3264850; PIDN:AAC24660.1; PID:g1617560; GSPDB:GN00
 A/Experimental source: strain MHOM/IL/81/Friedlin
 C/Genetics:
 A/Gene: LCPACAS5
 A/Map position: 1

Query Match 93.3%; Score 42; DB 2; Length 1607;
 Best Local Similarity 90.9%; Pred. No. 82;
 Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAK 11
 |||||
 Db 674 AAAAAAAAAAR 684

RESULT 13
 T31328
 fibroin - Chinese oak silkworm
 C/Species: Antherea pernyi (Chinese oak silkworm)
 C/Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
 C/Accession: T31328
 R/Suzuru, H.; Tamura, T.; Yukubiro, K.
 Submitted to the EMBL Data Library, August 1998
 A/Description: Characterization of the full length fibroin gene of a wild silkworm, Anth
 A/Reference number: Z20995
 A/Accession: T31328
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-2639 <SEZ>
 A/Cross-references: EMBL:AF083334; NID:g3450882; PID:g3450883; PIDN:AAC32606.1
 C/Genetics:
 A/Introns: 14/3

Query Match 93.3%; Score 42; DB 2; Length 2639;
 Best Local Similarity 90.9%; Pred. No. 1,2e+02;
 Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAK 11
 |||||
 Db 366 AAAAAAAAAAR 376

RESULT 14
 S58853
 homeotic protein ultrabithorax homolog - Junonia coenia (fragment)
 N/Alternate names: ultrabithorax homeodomain protein
 C/Species: Junonia coenia
 C/Date: 19-Mar-1997 #sequence_revision 29-Aug-1997 #text_change 15-Oct-1999
 C/Accession: S58853
 R/Warren, R.W.; Nagy, L.; Selegue, J.; Gates, J.; Carroll, S.
 Nature 372, 458-461, 1994
 A/Title: Evolution of homeotic gene regulation and function in flies and butterflies.
 A/Reference number: S58850; MUID:95075456; PMID:7840822
 A/Accession: S58853
 A/Status: nucleic acid sequence not shown; translation not shown
 A/Molecule type: mRNA
 A/Residues: 1-40 <WAR>
 A/Cross-references: EMBL:L42137
 A/Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1995
 C/Superfamily: unassigned homeobox proteins; homeobox homology
 C/Keywords: DNA binding; homeobox; nucleus; transcription regulation

Query Match 91.1%; Score 41; DB 2; Length 40;
 Best Local Similarity 90.9%; Pred. No. 7.2;
 Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAK 11
 |||||
 Db 25 AAAAAAAAAAQ 35

RESULT 15
 S02376
 antifreeze protein precursor - yellowtail flounder
 C/Species: Limanda ferruginea (yellowtail flounder)
 C/Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 24-Oct-2000
 C/Accession: S02376
 R/Scott, G.K.; Davies, P.L.; Shears, M.A.; Fletcher, G.L.
 Eur. J. Biochem. 168, 629-633, 1987
 A/Title: Structural variations in the alanine-rich antifreeze proteins of the Pleuronect
 A/Reference number: S02376; MUID:88029483; PMID:365937
 A/Accession: S02376
 A/Molecule type: mRNA
 A/Residues: 1-97 <SCO>
 A/Cross-references: EMBL:X06356; NID:g64041; PIDN:CAA29655.1; PID:g64042
 A/Note: part of this sequence, including the amino end of the mature protein, was confi
 C/Superfamily: antifreeze protein
 C/Keywords: antifreeze
 F:1-23/Domain: signal sequence #status predicted <SIG>
 F:24-48/Domain: propeptide #status predicted <PRO>
 F:49-96/Product: antifreeze protein #status predicted <MAT>

Query Match 91.1%; Score 41; DB 2; Length 97;
 Best Local Similarity 90.9%; Pred. No. 14;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAK 11
 |||||
 Db 57 AAAAAAAAAAK 67

Search completed: February 19, 2004, 10:05:03
 Job time : 23 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 19, 2004, 10:01:13 ; Search time 10 Seconds

(without alignments)
51.729 Million cell updates/sec

Title: US-09-551-336b-1
Perfect score: 45
Sequence: 1 AAAAAAAAAA 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues
Total number of hits satisfying chosen parameters: 127863

Minimum DB-seq length: 0
Maximum DB-seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	100.0	406	1 LHX2_HUMAN	P50458 homo sapien
2	45	100.0	406	1 LHX2_MOUSE	Q92082 mus musculu
3	45	100.0	435	1 ELM1_MOUSE	P10419 anthopleura
4	45	100.0	730	1 ELS_HUMAN	P15602 homo sapien
5	42	93.3	161	1 ID4_HUMAN	P47928 homo sapien
6	42	93.3	161	1 ID4_MOUSE	P41139 mus musculu
7	42	93.3	276	1 SX21_HUMAN	Q9Y651 mus musculu
8	42	93.3	280	1 SX21_CHICK	Q9W795 gallus gall
9	42	93.3	530	1 ZIC2_MOUSE	Q62520 mus musculu
10	42	93.3	764	1 CSM_DROVI	Q24708 drosophila
11	42	93.3	1077	1 HLB2_DROME	Q02308 drosophila
12	42	93.3	1533	1 PUM_DROME	P25822 drosophila
13	41	91.1	97	1 ANP_LIMFE	P09031 limanda fer
14	41	91.1	231	1 ASH1_MOUSE	Q02067 mus musculu
15	41	91.1	233	1 ASH1_RAT	P19359 rattus norv
16	41	91.1	236	1 ASH1_HUMAN	P50853 homo sapien
17	41	91.1	314	1 PMXB_HUMAN	Q99453 homo sapien
18	41	91.1	314	1 PMXB_MOUSE	Q35690 mus musculu
19	41	91.1	323	1 HXDB_MOUSE	P23413 mus musculu
20	41	91.1	337	1 HAIR_DROME	Q10723 drosophila
21	41	91.1	338	1 HXDB_HUMAN	P31377 homo sapien
22	41	91.1	392	1 HME1_HUMAN	Q05925 homo sapien
23	41	91.1	437	1 ZP12_MOUSE	P56224 brachydantio
24	41	91.1	441	1 P033_BRARE	Q90436 brachydantio
25	41	91.1	451	1 P031_RAT	P20267 rattus norv
26	41	91.1	507	1 IRX3_MOUSE	P81067 mus musculu
27	41	91.1	649	1 ARA_VOLCA	Q10723 volvox cart
28	41	91.1	701	1 TBX2_MOUSE	Q60707 mus musculu
29	41	91.1	702	1 TBX2_HUMAN	Q13207 homo sapien
30	41	91.1	747	1 ELS_BOVIN	P04385 bos taurus
31	41	91.1	860	1 ELS_MOUSE	P54320 mus musculu
32	41	91.1	864	1 ELS_RAT	Q99372 rattus norv
33	41	91.1	880	1 BRCT_DROME	Q24206 drosophila

34	41	91.1	1067	1 BAB2_DROME	Q9W0K4 drosophila
35	41	91.1	1355	1 SALM_DROME	P39770 drosophila
36	41	91.1	1402	1 SALM_DROVI	P39806 drosophila
37	41	91.1	1596	1 MAM_DROME	P21519 drosophila
38	41	91.1	1669	1 ASX_DROME	Q9V727 drosophila
39	41	91.1	1783	1 RAA3_CHLE	Q9TEC4 chlamydomon
40	40	88.9	109	1 RLAI_TRYCR	P26643 trypanosoma
41	40	88.9	217	1 HAN2_HUMAN	Q95300 homo sapien
42	40	88.9	289	1 HXD8_MOUSE	P23463 mus musculu
43	40	88.9	297	1 MBL_DROME	Q16011 drosophila
44	40	88.9	304	1 GSH2_HUMAN	Q9BZM3 homo sapien
45	40	88.9	305	1 GSH2_MOUSE	P31316 mus musculu

ALIGNMENTS

RESULT 1	ID	Sequence	STANDARD	PRT	406 AA
AC	P50458	Q95860			
DT	01-OCT-1996	(Rel. 34, Created)			
DT	30-MAY-2000	(Rel. 39, Last sequence update)			
DT	28-FEB-2003	(Rel. 41, Last annotation update)			
DE	LIM/homeobox protein Lhx2 (Homeobox protein Lhx2)				
GN	LHX2 OR LHX2				
OS	Homo sapiens (Human)				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
OX	MEBL_Taxid=9606;				
RN	(1)				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Placenta;				
RX	MEDLINE=96226351; PubMed=8649822;				
RA	Wu H.-K., Heng H.H.Q., Siderovski D.P., Dong W.-F., Okuno Y.,				
RA	Shi X.-M., Tsui L.-C., Minden M.D.;				
RT	"Identification of a human LIM-Hox gene, hLH-2, aberrantly expressed				
RT	in chronic myelogenous leukaemia and located on 9q33-34.1.";				
RL	entogene 12:1205-1212 (1996).				
RN	(12)				
RP	SEQUENCE OF 18-406 FROM N.A.				
RX	MEDLINE=99162575; PubMed=10051612;				
RA	Rincon-Limas D.E., Lu C.-H., Calleja M.,				
RA	Rodriguez-Esteban C., Izpisua-Belmonte J.C., Bozas J.;				
RT	"Conservation of the expression and function of apertous orthologs in				
RT	drosophila and mammals.";				
CC	Proc. Natl. Acad. Sci. U.S.A. 96:2165-2170 (1999).				
CC	-1- FUNCTION: TRANSCRIPTIONAL REGULATORY PROTEIN INVOLVED IN THE				
CC	CONTROL OF CELL DIFFERENTIATION IN DEVELOPING LYMPHOID AND				
CC	NEURAL CELL TYPES (BY SIMILARITY).				
CC	-1- SUBCELLULAR LOCATION: Nuclear (Probable).				
CC	-1- SIMILARITY: Contains 1 homeobox domain.				
CC	-1- SIMILARITY: Contains 2 LIM zinc-binding domains.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@sib-sib.ch).				
CC	-----				
CC	EMBL: U11701; AAB08752.1; -				
CC	EMBL: AF124735; AAD20013.1; -				
CC	HSSP: P06601; 1FJL.				
CC	GeneW: HGNC:6594; LHX2.				
CC	MIM: 603759; -				
CC	GO: GO:0007048; P:oncogenesis; TAS.				
CC	InterPro: IPR001356; Homeobox.				
CC	InterPro: IPR001781; LIM.				
CC	InterPro: IPR007107; LIM_homo.				
CC	Pfam: PF00046; homeobox; 1.				
CC	Pfam: PF00412; LIM; 2.				

DR Prodom; PD000010; Homeobox; 1.
 DR Prodom; PD000094; LIM; 2.
 DR SMART; SM00389; HOX; 1.
 DR SMART; SM00132; LIM; 2.
 DR PROSITE; PS00478; LIM DOMAIN 1; 2.
 DR PROSITE; PS50023; LIM DOMAIN 2; 2.
 DR PROSITE; PS50027; HOMEBOX 1; 1.
 DR PROSITE; PS50071; HOMEBOX 2; 1.
 KM Homeobox; DNA-binding; Zinc; Transcription regulation.
 FT DOMAIN 1.
 FT DOMAIN 115 105
 FT DOMAIN 168
 FT DOMAIN 187 196
 FT DNA BIND 266 325
 FT DOMAIN 307 323
 FT CONFLICT 22 28
 FT CONFLICT 54 54
 FT CONFLICT 187 187
 FT CONFLICT 190 192
 FT CONFLICT 196 196
 FT CONFLICT 241 242
 FT CONFLICT 391 406
 SQ SEQUENCE 406 AA; 44373 MW; DF73AAC81867D30D CRC64;
 Query Match 100.0%; Score 45; DB 1; Length 406;
 Best Local Similarity 100.0%; Pred. No. 4.6; Mismatches 0; Indels 0; Gaps 0;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAA 11
 Db 187 AAAAAAAAAA 197

RESULT 2
 LHX2 MOUSE STANDARD; PRT; 406 AA.
 ID LHX2 MOUSE
 AC 092052;
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE LIM/homeobox protein lhx2.
 GN LHX2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99162575; PubMed=10051612;
 RA Rincon-Limas D.E., Lu C.-H., Canal I., Calleja M.,
 RA Rodriguez-Becben C., Izpisua-Belmonte J.C., Botos J.;
 RT "Conservation of the expression and function of apterous orthologs in
 RT Drosophila and mammals.";
 RL Proc. Natl. Acad. Sci. U.S.A. 96:2165-2170(1999).
 CC -1- FUNCTION: TRANSCRIPTIONAL REGULATOR PROTEIN INVOLVED IN THE
 CC CONTROL OF CELL DIFFERENTIATION IN DEVELOPING LYMPHOID AND
 CC NEURAL CELL TYPES (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Nuclear (probable).
 CC -1- SIMILARITY: Contains 1 homeobox domain.
 CC -1- SIMILARITY: Contains 2 LIM zinc-binding domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF124734; AAD20012.1; -
 CC HSSP; P06601; 1FJL.
 CC TRANSFAC; T01969; -

DR MGD; MGI:96785; Lhx2.
 DR GO; GO:0007498; P:neuroderm development; IMP.
 DR GO; GO:0007399; P:neurogenesis; IMP.
 DR InterPro; IPR001356; Homeobox.
 DR InterPro; IPR001781; LIM.
 DR InterPro; IPR007107; LIM_homeo.
 DR Pfam; PF00412; LIM; 2.
 DR Prodom; PD000010; Homeobox; 1.
 DR Prodom; PD000094; LIM; 2.
 DR SMART; SM00389; HOX; 1.
 DR SMART; SM00132; LIM; 2.
 DR PROSITE; PS00478; LIM DOMAIN 1; 2.
 DR PROSITE; PS50023; LIM DOMAIN 2; 2.
 DR PROSITE; PS50027; HOMEBOX 1; 1.
 DR PROSITE; PS50071; HOMEBOX 2; 1.
 KM Homeobox; DNA-binding; Zinc; Transcription regulation.
 FT DOMAIN 1.
 FT DOMAIN 53 105
 FT DOMAIN 115 168
 FT DOMAIN 187 196
 FT DNA BIND 266 325
 FT DOMAIN 307 323
 FT CONFLICT 22 28
 FT CONFLICT 54 54
 FT CONFLICT 187 187
 FT CONFLICT 190 192
 FT CONFLICT 196 196
 FT CONFLICT 241 242
 FT CONFLICT 391 406
 SQ SEQUENCE 406 AA; 44419 MW; FE7B4E76454D6A90 CRC64;
 Query Match 100.0%; Score 45; DB 1; Length 406;
 Best Local Similarity 100.0%; Pred. No. 4.6; Mismatches 0; Indels 0; Gaps 0;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAA 11
 Db 187 AAAAAAAAAA 197

RESULT 3
 FMRI ANTEN STANDARD; PRT; 435 AA.
 ID FMRI ANTEN
 AC P10419;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Antio-Ramide neuropeptides type 1 precursor.
 OS Anthopleura elegantissima (Sea anemone).
 OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actiniaria;
 OC Nyanthaeae; Actiniidae; Anthopleura.
 OX NCBI_TaxID=6110;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93054550; PubMed=1429603;
 RA Schmittzler C., Daxner D., Diekhoff D., Grimmelikhuijzen C.J.P.;
 RT "Identification of a novel type of processing sites in the precursor
 RT for the sea anemone neuropeptide Antio-Ramide (<Glu-Gly-Arg-Phe-NH2>
 RT from Anthopleura elegantissima.";
 RL J. Biol. Chem. 267:22534-22541 (1992).
 CC [2]
 RP PARTIAL SEQUENCE (ANTHO-RFAMIDE).
 RX MEDLINE=87092339; PubMed=2879289;
 RA Grimmelikhuijzen C.J.P., Graff D.;
 RT "Isolation of PyroGlu-Gly-Arg-Phe-NH2 (Antio-Ramide), a neuropeptide
 RT from sea anemones.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:9817-9821(1986).
 CC -1- FUNCTION: NOT KNOWN BUT IT COULD ACT AS A TRANSMITTER AT
 CC NEURONMUSCULAR SYNAPSES.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; M98269; AAA27738.1; -.
 DR PIR; A44308; A44308.
 DR InterPro; IPR002544; FARP.
 DR Pfam; PF01581; FARP; 13.
 DR Neuropeptide; Amidation; Repeat; Signal; Pyrrolidone carboxylic acid.
 FT SIGNAL 1 22
 FT PEPTIDE 194 197 ANTHO-RFAMIDE.
 FT PEPTIDE 202 205 ANTHO-RFAMIDE.
 FT PEPTIDE 210 213 ANTHO-RFAMIDE.
 FT PEPTIDE 218 221 ANTHO-RFAMIDE.
 FT PEPTIDE 226 229 ANTHO-RFAMIDE.
 FT PEPTIDE 234 237 ANTHO-RFAMIDE.
 FT PEPTIDE 242 245 ANTHO-RFAMIDE.
 FT PEPTIDE 250 253 ANTHO-RFAMIDE.
 FT PEPTIDE 258 261 ANTHO-RFAMIDE.
 FT PEPTIDE 266 269 ANTHO-RFAMIDE.
 FT PEPTIDE 274 277 ANTHO-RFAMIDE.
 FT PEPTIDE 282 285 ANTHO-RFAMIDE.
 FT PEPTIDE 290 293 ANTHO-RFAMIDE.
 FT PEPTIDE 298 301 ANTHO-RFAMIDE.
 FT PEPTIDE 306 309 ANTHO-RFAMIDE.
 FT PEPTIDE 314 317 ANTHO-RFAMIDE.
 FT PEPTIDE 322 325 ANTHO-RFAMIDE.
 FT PEPTIDE 330 333 ANTHO-RFAMIDE.
 FT PEPTIDE 343 346 ANTHO-RFAMIDE.
 FT PEPTIDE 356 359 ANTHO-RFAMIDE.
 FT PEPTIDE 369 372 ANTHO-RFAMIDE.
 FT DOMAIN 376 386 POLY-ALA.
 FT MOD_RES 194 194 PYRROLIDONE CARBOXYLIC ACID.
 FT MOD_RES 197 197 AMIDATION (G-198 PROVIDE AMIDE GROUP).
 FT MOD_RES 205 205 AMIDATION (G-206 PROVIDE AMIDE GROUP).
 FT MOD_RES 213 213 AMIDATION (G-214 PROVIDE AMIDE GROUP).
 FT MOD_RES 221 221 AMIDATION (G-222 PROVIDE AMIDE GROUP).
 FT MOD_RES 229 229 AMIDATION (G-230 PROVIDE AMIDE GROUP).
 FT MOD_RES 237 237 AMIDATION (G-238 PROVIDE AMIDE GROUP).
 FT MOD_RES 245 245 AMIDATION (G-246 PROVIDE AMIDE GROUP).
 FT MOD_RES 253 253 AMIDATION (G-254 PROVIDE AMIDE GROUP).
 FT MOD_RES 261 261 AMIDATION (G-262 PROVIDE AMIDE GROUP).
 FT MOD_RES 269 269 AMIDATION (G-270 PROVIDE AMIDE GROUP).
 FT MOD_RES 277 277 AMIDATION (G-278 PROVIDE AMIDE GROUP).
 FT MOD_RES 285 285 AMIDATION (G-286 PROVIDE AMIDE GROUP).
 FT MOD_RES 293 293 AMIDATION (G-294 PROVIDE AMIDE GROUP).
 FT MOD_RES 301 301 AMIDATION (G-302 PROVIDE AMIDE GROUP).
 FT MOD_RES 309 309 AMIDATION (G-310 PROVIDE AMIDE GROUP).
 FT MOD_RES 317 317 AMIDATION (G-318 PROVIDE AMIDE GROUP).
 FT MOD_RES 325 325 AMIDATION (G-326 PROVIDE AMIDE GROUP).
 FT MOD_RES 333 333 AMIDATION (G-334 PROVIDE AMIDE GROUP).
 FT MOD_RES 346 346 AMIDATION (G-347 PROVIDE AMIDE GROUP).
 FT MOD_RES 359 359 AMIDATION (G-360 PROVIDE AMIDE GROUP).
 FT MOD_RES 372 372 AMIDATION (G-373 PROVIDE AMIDE GROUP).
 SO SEQUENCE 435 AA; 50940 MW; B0C44020CD58061 CRC64;

Query Match 100.0%; Score 45; DB 1; Length 435;
 Best Local Similarity 100.0%; Pred. No. 4.8;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAK 11
 DB 377 AAAAAAAAAAAK 387

RESULT 4
 ELS_HUMAN STANDARD; PRT; 730 AA.
 AC P15502; Q14233; Q14238;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Elastin precursor (Tropoelastin).
 GN ELN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxId=9606;
 RN (1)
 RP SEQUENCE FROM N.A. (ISOFORM B).
 RX MEDLINE=87289668; PubMed=3039501;
 RA Indik Z., Yeh H., Ornstein-Goldstein N., Sheppard P., Anderson N.,
 RA Rosenbloom J.C., Peltonen L., Rosenbloom J.;
 RA "Alternative splicing of human elastin mRNA indicated by sequence
 analysis of cloned genomic and complementary DNA."
 Proc. Natl. Acad. Sci. U.S.A. 84:5680-5684(1987).
 RN (2)
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Skin fibroblast;
 RX MEDLINE=89009960; PubMed=3171221;
 RA Fazio M.J., Olsen D.R., Kaub E.A., Baldwin C.T., Indik Z.,
 RA Ornstein-Goldstein N., Yeh H., Rosenbloom J., Ultio J.;
 RA "Cloning of full-length elastin cDNAs from a human skin fibroblast
 recombinant cDNA library: further elucidation of alternative splicing
 utilizing exon-specific oligonucleotides."
 Invest. Dermatol. 91:458-464(1988).
 RN (3)
 RP SEQUENCE OF 164-724 FROM N.A. (ISOFORM B).
 RC TISSUE=Placenta;
 RX MEDLINE=88156138; PubMed=2831431;
 RA Fazio M.J., Olsen D.R., Kuivaniemi H., Chu M.L., Davidson J.M.,
 RA Rosenbloom J., Ultio J.;
 RA "Isolation and characterization of human elastin cDNAs, and age-
 associated variation in elastin gene expression in cultured skin
 fibroblasts."
 Invest. 58:270-277(1988).
 RN (4)
 RP SEQUENCE OF 603-730 FROM N.A.
 RC TISSUE=Hippocampus, and placenta;
 RX MEDLINE=96291399; PubMed=8689688;
 RA Frangiskakis J.M., Earle A.K., Morris C.A., Mervis C.B.,
 RA Bertrand J., Robinson B.F., Klein B.P., Ensing G.J., Everett L.A.,
 RA Green E.D., Proeschel C., Gutowski N.J., Noble M., Atkinson D.L.,
 RA Odenberg S.J., Keating M.T.;
 RA "Lim-kinase1 hemizygosity implicated in impaired visuospatial
 constructive cognition."
 Cell 86:59-69(1996).
 RL Cell 86:59-69(1996).
 CC -1- FUNCTION: MATRIX STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND
 CC NUCAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.
 CC -1- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER
 CC INTO AN EXTENSIBLE 3D NETWORK.
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Comment=Additional isoforms seem to exist;
 CC Name=1;
 CC IsoId=P15502-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=P15502-2; Sequence=VSP_004243;
 CC -1- PTM: THE CROSSLINKS ARE MADE OF DEAMINATED LYS.
 CC -1- DISEASE: Haploinsufficiency of ELN may be the cause of certain
 CC cardiovascular and musculo-skeletal abnormalities observed in
 CC Williams-Beuren syndrome (WBS), a rare developmental disorder. It
 CC is a contiguous gene deletion syndrome involving genes from
 CC chromosome band 7q11.23.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by, and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M17282; AAC98394.1; -;
 CC EMBL; M16983; AAC98394.1; JOINED.
 CC EMBL; M17265; AAC98394.1; JOINED.
 CC EMBL; M17266; AAC98394.1; JOINED.
 CC EMBL; M17267; AAC98394.1; JOINED.
 CC EMBL; M17268; AAC98394.1; JOINED.

```

DR EMBL: M17270: AAC98394.1; JOINED.
DR EMBL: M17271: AAC98394.1; JOINED.
DR EMBL: M17272: AAC98394.1; JOINED.
DR EMBL: M17273: AAC98394.1; JOINED.
DR EMBL: M17275: AAC98394.1; JOINED.
DR EMBL: M17276: AAC98394.1; JOINED.
DR EMBL: M17277: AAC98394.1; JOINED.
DR EMBL: M17278: AAC98394.1; JOINED.
DR EMBL: M17279: AAC98394.1; JOINED.
DR EMBL: M17280: AAC98394.1; JOINED.
DR EMBL: M17281: AAC98394.1; JOINED.
DR EMBL: M36860: AAAS2382.1; -.
DR EMBL: M24782: AAAS3190.1; -.
DR EMBL: U62292: AAB17544.1; -.
DR EMBL: X15603: CA835627.1; -.
DR PIR: A32707; EATH.
DR HSSP: P50099; 1ZFU.
DR Genew; HGNC:3327; ELN.
DR MIM: 130160; -.
DR MIM: 194050; -.
DR GO: GO:0005578; C:extracellular matrix; TAS.
DR GO: GO:0005615; C:extracellular space; TAS.
DR GO: GO:0005201; F:extracellular matrix structural constituent; TAS.
DR GO: GO:0008283; F:cell proliferation; TAS.
DR GO: GO:0008015; P:circulation; TAS.
DR GO: GO:0007397; P:histogenesis and organogenesis; TAS.
DR GO: GO:0007385; P:respiratory gaseous exchange; TAS.
DR InterPro: IPR003979; tropoelastin.
DR PRINTS: PR01500; TROPOELASTIN.
KW Structural protein; Connective tissue; Repeat; Signal;
  Williams-Beuren syndrome; Alternative splicing.
FT SIGNAL 1 26
FT CHAIN 27 730 ELASTIN.
FT DISULFID 720 725 BY SIMILARITY.
FT VAASPLIC 472 477 Missing (in isoform 2).
  /FTid=VSP_004243.
SQ SEQUENCE 730 AA; 63260 MW; AB060D5BA567AE46 CRC64;

Query Match 100.0%; Score 45; DB 1; Length 730;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAK 11
Db 302 AAAAAAAAAAK 312

RESULT 5
ID4_HUMAN
AC P47928; Q13005; STANDARD; PRT; 161 AA.
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE DNA-binding protein inhibitor ID-4.
GN ID4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Abdominal adipose tissue;
RA Kieseling T.L., Christy B.A.;
RL Submitted (JUN-1995) to the EMBL/Genbank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC MEDLINE=99087490; PubMed=9872455;
RA Rigole M., Rich T., Gross-Morand M.S., Molina-Gomes D.,
RA Viegas-Pequignot E., Junien C.;
RT "cDNA cloning, tissue distribution and chromosomal localization of
  the human ID4 gene.";
RL DNA Res. 5:309-313(1998).

```

FN	[3]	SEQUENCE FROM N.A.
RP		MEDLINE=95394461; PubMed=7665172;
RX		Paglicca A., Bartoli P.C., Saccone S., della Valle G., Iania L.;
RA		"Molecular cloning of ID4, a novel dominant negative helix-loop-helix
RT		human gene on chromosome 6p21.3-p22.";
RL		Genomics 27:200-203(1995).
RN	[4]	SEQUENCE FROM N.A.
RP		Mashreghi-Mohammadi M.;
RA		Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN	[5]	SEQUENCE FROM N.A.
RP		TISSUE=Uterus;
RC		MEDLINE=22388257; PubMed=12477932;
RX		Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA		Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA		Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA		Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heist F.,
RA		Datchenko L., Marusik K., Farmer A.A., Rubin G.M., Hong L.,
RA		Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.E.,
RA		Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA		Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA		Boeak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA		Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA		Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA		Faney J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA		Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA		Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA		Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA		Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
RA		Scherech A., Schein J.E., Jones S.J.M., Marra M.A.;
RT		"Generation and initial analysis of more than 15,000 full-length
RT		human and mouse cDNA sequences."
RL		Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC	-1-	FUNCTION: ID (INHIBITOR OF DNA BINDING) HLH PROTEINS LACK A BASIC
CC	-1-	DNA-BINDING DOMAIN BUT ARE ABLE TO FORM HETERODIMERS WITH OTHER
CC	-1-	HLH PROTEINS. THEREBY INHIBITING DNA BINDING.
CC	-1-	SUBUNIT: HETERODIMER WITH OTHER HLH PROTEINS.
CC	-1-	SUBCELLULAR LOCATION: Nucleus.
CC	-1-	SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC	-1-	TRANSCRIPTION FACTORS. "ID" SUBFAMILY.
CC		-----
CC		This SWISS-PROT entry is copyright. It is produced through a collaboration
CC		between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC		the European Bioinformatics Institute. There are no restrictions on its
CC		use by non-profit institutions as long as its content is in no way
CC		modified and this statement is not removed. Usage by and for commercial
CC		entities requires a license agreement (See http://www.isb.ch/announce/
CC		or send an email to license@isb-sib.ch).
CC		-----
DR	EMBL	U28368; AAA73923.1; -
DR	EMBL	Y07958; CAA69255.1; -
DR	EMBL	U16153; AAA82882.1; -
DR	EMBL	AL022726; CAA18779.1; -
DR	EMBL	BC014941; AAH14941.1; -
DR	PIR	G01855; G01855.
DR	GeneW	HGNC:5363; ID4.
DR	MIM	600581; -
DR	GO	GO:0003714; F:transcription co-repressor activity; TAS.
DR	GO	GO:0006357; P:regulation of transcription from Pol II pro. .; TAS.
DR	InterPro	IPR001092; HLH_Basic.
DR	Pfam	PF00010; HLH_1.
DR	SMART	SM00353; HLH_1.
DR	PROSITE	PS00038; HLH_1; 1.
DR	PROSITE	PS50888; HLH_2; 1.
DR		Nuclear protein.
KM	DOMAIN	39 48 POLY-ALA.
FT	DOMAIN	65 105 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT	DOMAIN	118 124 POLY-PRO.
FT	CONFLICT	10 14 SGRA -> RPLR (IN REF. 3).
FT	CONFLICT	39 40 AA -> Q (IN REF. 3).
FT	CONFLICT	77 79 RLV -> WL (IN REF. 3).

SQ SEQUENCE 161 AA; 16622 MW; 5B14847AE7337339 CRC64;
 Query Match 93.3%; Score 42; DB 1; Length 161;
 Best Local Similarity 90.9%; Pred. No. 6;
 Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAA 11
 DB 39 AAAAAAAAAA 49
 RESULT 6
 ID4_MOUSE STANDARD; PRT; 161 AA.
 AC P4139;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE DNA-binding protein inhibitor ID-4.
 ID4 OR ID-4 OR ID84.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 NCBI_TaxID=10090;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BA1B/C; TISSUE=Bone marrow;
 RX MEDLINE=94188125; PubMed=8139914;
 RA Riechmann V., van Crecchten I., Sablitzky F.;
 RT "The expression pattern of Id4, a novel dominant negative helix-loop-
 RT helix protein, is distinct from Id1, Id2 and Id3."
 RL Nucleic Acids Res. 22:749-755(1994).
 [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99057583; PubMed=9838043;
 RA van Crecchten I., Cinato E., Fox M., King E.R., Newton J.S.,
 RA Riechmann V., Sablitzky F.;
 RT "Structure, chromosomal localization and expression of the murine
 RT dominant negative helix-loop-helix Id4 gene."
 RL Biochim. Biophys. Acta 1443:55-64(1998).
 [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99051333; PubMed=9831657;
 RA Maniani A., Hernandez M.C., Kuo W.-L., Israel M.A.;
 RT "The mouse Id2 and Id4 genes: structural organization and chromosomal
 RT localization."
 RL Gene 222:229-235(1998).
 CC -1- FUNCTION: ID (INHIBITOR OF DNA BINDING) HLH PROTEINS LACK A BASIC
 CC DNA-BINDING DOMAIN BUT ARE ABLE TO FORM HETERODIMERS WITH OTHER
 CC HLH PROTEINS, THEREBY INHIBITING DNA BINDING.
 CC -1- SUBUNIT: HETERODIMER WITH OTHER HLH PROTEINS.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLL) FAMILY OF
 CC TRANSCRIPTION FACTORS. "ID" SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X75018; CAAS2926.1; -;
 DR EMBL; AJ001972; CA05120.1; -;
 DR EMBL; AF077859; AAD05213.1; -;
 DR PIR; S43260; S43260.
 DR TRANSFAC; T01658; -;
 DR MGD; MGI:99414; Id4.
 DR InterPro; IPR001092; HLH_basic.
 DR Pfam; PFO0010; HLH; 1.
 DR SMART; SM00353; HLH; 1.
 DR PROSITE; PS00038; HLH_1; 1.

DR PROSITE; PS50888; HLH_2; 1.
 KW Nuclear protein.
 FT DOMAIN 39 48 POLY-ALA.
 FT DOMAIN 65 105 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
 FT DOMAIN 118 124 POLY-PRO.
 SQ SEQUENCE 161 AA; 16596 MW; 2DCFA7AF7EE7EED CRC64;
 Query Match 93.3%; Score 42; DB 1; Length 161;
 Best Local Similarity 90.9%; Pred. No. 6;
 Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAA 11
 DB 39 AAAAAAAAAA 49
 RESULT 7
 SX21_HUMAN STANDARD; PRT; 276 AA.
 AC Q9Y651;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Transcription factor SOX-21.
 GN SOX21.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NCBI_TaxID=9606;
 [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99373261; PubMed=10441749;
 RA Malas S., Duthe S., Deloukas P., Episkopou V.;
 RT "The isolation and high-resolution chromosome mapping of human SOX14
 RT and SOX21, two members of the SOX gene family related to SOX1, SOX2,
 RL Mamm. Genome 10:934-937(1999).
 CC -1- FUNCTION: ACTS AS A NEGATIVE REGULATOR OF TRANSCRIPTION (BY
 CC SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -1- SIMILARITY: Contains 1 HMG box domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF107044; AAC95381.1; -;
 DR HSP; Q05066; IHR1.
 DR TRANSFAC; T04921; -;
 DR GeneW; HGNC:11197; SOX21.
 DR MIM; 604974; -;
 DR GO; GO:0003702; F-RNA polymerase II transcription factor acti. . .; TAS.
 DR GO; GO:0006357; P-regulation of transcription from Pol II pro. . .; TAS.
 DR InterPro; IPR000910; HMG_12_box.
 DR Pfam; PF00505; HMG_box; 1.
 DR SMART; SM00398; HMG; 1.
 DR PROSITE; PS50118; HMG_BOX_2; 1.
 KW Transcription regulation; Repressor; DNA-binding; Nuclear protein.
 FT DNA BIND 8 76 HMG_BOX.
 SQ SEQUENCE 276 AA; 28580 MW; 99CC89B7BC9A6B CRC64;
 Query Match 93.3%; Score 42; DB 1; Length 276;
 Best Local Similarity 90.9%; Pred. No. 8; 9;
 Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAA 11
 DB 137 AAAAAAAAAA 147

RESULT 8
 ID SKX1_CHICK STANDARD; PRT; 280 AA.
 AC Q9W7R5;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Transcription factor SOX-21.
 GN SOX21.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OC NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
 RC TISSUE=Liver;
 RX MEDLINE=99400103; Pubmed=10473124;
 RA Uchikawa M., Kamachi Y., Kondoh H.;
 RT "Two distinct subgroups of Group B Sox genes for transcriptional
 RT activators and repressors: their expression during embryonic
 RT organogenesis of the chicken.";
 RL Mech. Dev. 84:103-120(1999).
 CC -1- FUNCTION: Acts as a negative regulator of transcription.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -1- TISSUE SPECIFICITY: Expressed predominantly in CNS.
 CC -1- SIMILARITY: Contains 1 HMG box domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AB026623; BAA7266.1; -;
 DR HSSP; Q05066; 1HRV;
 DR InterPro; IPR000910; HMG_12_box.
 DR Pfam; PF00505; HMG_box; 1.
 DR SMART; SM00398; HMG; 1.
 DR PROSITE; PSS0118; HMG_BOX_2; 1.
 DR Transcription regulation; Repressor; DNA-binding; Nuclear protein.
 FT DNA BIND 8 76 HMG_BOX.
 FT DOMAIN 137 146 POLY-ALA.
 FT DOMAIN 154 166 POLY-ALA.
 FT DOMAIN 183 187 POLY-SER.
 FT DOMAIN 212 223 POLY-ALA.
 SQ SEQUENCE 280 AA; 28796 MW; DA858428B29F1F47 CRC64;
 Query Match 93.3%; Score 42; DB 1; Length 280;
 Best Local Similarity 90.9%; Pred. No. 9;
 Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAA 11
 Db 137 AAAAAAAAAA 147

RESULT 9
 ID ZIC2_MOUSE STANDARD; PRT; 530 AA.
 AC Q62520;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Zinc finger protein ZIC 2 (Zinc finger protein of the cerebellum 2).
 GN ZIC2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Cerebellum;
 RX MEDLINE=96132643; Pubmed=8557628;
 RA Aruga J., Nagai T., Tokuyama T., Hayashizaki Y., Okazaki Y.,
 RA Chapman V.M., Mikoshiba K.;
 RT "The mouse zic gene family. Homologues of the Drosophila pair-rule
 RT gene odd-paired.";
 RL J. Biol. Chem. 271:1043-1047(1996).
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- TISSUE SPECIFICITY: CNS. A HIGH LEVEL EXPRESSION IS SEEN IN THE
 CC CEREBELLUM.
 CC -1- SIMILARITY: BELONGS TO THE GLI FAMILY OF C2H2-TYPE ZINC-FINGER
 CC PROTEINS.
 CC -1- SIMILARITY: Contains 4 C2H2-type zinc fingers.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D70848; BAA1115.1; -;
 DR HSSP; P08047; 1SP2.
 DR TRANSFAC; T04670; -;
 DR MGD; MGI:106679; Zic2.
 DR GO; GO:0007417; P; Central nervous system development; IGI.
 DR InterPro; IPR007087; ZnF_C2H2.
 DR Pfam; PF00096; Zf_C2H2; 4.
 DR Prodom; PD000003; ZnF_C2H2; 1.
 DR SMART; SM00355; ZnF_C2H2; 4.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
 DR PROSITE; PSS0157; ZINC_FINGER_C2H2_2; 4.
 DR PROSITE; PSS0157; ZINC_FINGER_C2H2_2; 4.
 DR Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein.
 FT DOMAIN 20 23 POLY-HIS.
 FT DOMAIN 25 33 POLY-ALA.
 FT DOMAIN 89 97 POLY-ALA.
 FT DOMAIN 227 231 POLY-ALA.
 FT DOMAIN 232 239 POLY-HIS.
 FT ZN_FING 300 327 C2H2-TYPE 1 (ATYPICAL).
 FT ZN_FING 333 357 C2H2-TYPE 2.
 FT ZN_FING 363 387 C2H2-TYPE 3.
 FT ZN_FING 393 415 C2H2-TYPE 4.
 FT ZN_FING 456 470 POLY-ALA.
 FT DOMAIN 501 512 POLY-ALA.
 SQ SEQUENCE 530 AA; 55492 MW; 0065BD75B52E7D2 CRC64;
 Query Match 93.3%; Score 42; DB 1; Length 530;
 Best Local Similarity 90.9%; Pred. No. 14;
 Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAA 11
 Db 458 AAAAAAAAAA 468

RESULT 10
 ID CSW_DROVI STANDARD; PRT; 764 AA.
 AC Q24708;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Protein-tyrosine phosphatase corkscrew (EC 3.1.3.48) (Fragment).
 GN CSW.
 OS Drosophila virilis (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OC NCBI_TaxID=7244;
 RN [1]

RP SEQUENCE FROM N.A.
 RA Melnick M.B., Melnick C.B., Larsen I., Perlmom N., Perkins L.A.;
 RT "The role of the Drosophila corkscrew protein as a transducer
 RL downstream of receptor tyrosine kinases is functionally conserved.";
 CC Submitted (MAR-1995) to the EMBL/Genbank/DBD database.
 CC -1- FUNCTION: Required in all receptor tyrosine kinase signaling
 CC pathways. Functions downstream of the receptor tyrosine kinase
 CC torso, acting in concert with D-Raf via tailless. Also functions
 CC downstream of Egfr (epidermal growth factor receptor) and btl
 CC (fibroblast growth factor receptor). The SH2 domain suggests that
 CC csw effects its role by mediating heteromeric protein
 CC interactions. Maternally required for normal determination of cell
 CC fates at the termini of the embryo. Required for cell fate
 CC specification of the ventral ectoderm. In the developing embryonic
 CC CNS and for embryonic tracheal cell migration. Functions during
 CC imaginal development for proper formation of adult structures such
 CC as eyes, arista, L5 wing vein and the tarsal claw (By
 CC similarity).
 CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein
 CC tyrosine + phosphate.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
 CC TYROSINE PHOSPHATASE FAMILY. SUBCLASS THAT CONTAINS SH2 DOMAINS.
 CC -1- SIMILARITY: Contains at least 1 SH2 domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U23356; AAB02545.1; --
 CC HSSP: Q06124; ZSPH.
 CC DR F1ybase; F8gn0016496; Dvir\csw.
 CC DR InterPro; IPR000980; SH2.
 CC DR InterPro; IPR000387; TYR_phosphatase.
 CC DR InterPro; IPR000242; Tyr_PP.
 CC DR Pfam; PF001017; SH2; 1.
 CC DR Pfam; PF00102; Y_phosphatase; 1.
 CC DR Pfam; PR00700; PRTYPHPTASE.
 CC DR PRINTS; PR00401; SH2DOMAIN.
 CC DR ProDom; PD000093; SH2; 1.
 CC DR SMART; SM00194; PTPc; 1.
 CC DR SMART; SM00252; SH2; 1.
 CC DR PROSITE; PS50001; SH2; 1.
 CC DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
 CC DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
 CC DR PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 1.
 CC DR Hydrolyase; SH2 domain; Developmental protein.
 CC FT NON_TER 1 1
 CC FT DOMAIN 1 95
 CC FT DOMAIN 117 522 SH2.
 CC FT DOMAIN 174 325 PROTEIN-TYROSINE PHOSPHATASE.
 CC FT ACT_SITE 460 460 PTPASE INSERT (CYS/SER-RICH).
 CC FT DOMAIN 186 191 BY SIMILARITY.
 CC FT DOMAIN 278 281 POLY-SER.
 CC FT DOMAIN 576 600 POLY-ALA.
 CC FT DOMAIN 613 616 ALA-RICH.
 CC FT DOMAIN 617 656 POLY-ASN.
 CC FT DOMAIN 687 690 SER-RICH.
 CC FT DOMAIN 694 753 POLY-GLN.
 CC FT DOMAIN 753 761 ALA-RICH.
 CC FT DOMAIN 761 761 POLY-PRO.
 CC SQ SEQUENCE 764 AA; 82125 MW; 66008EAB560A2F7D CRC64;
 QY Query Match 93.3%; Score 42; DB 1; Length 764;
 DB Best local Similarity 90.9%; Pred. No. 19;
 Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 1 AAAAAAAAAAANK 11
 :|||||
 719 SAAAAAANK 729

RESULT 11
 HLES_DROME STANDARD; PRT: 1077 AA.
 ID HLES_DROME Q02308; Q9VDK0; Q9VDK1;
 AC Q02308; Q9VDK0; Q9VDK1;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Hairless protein.
 GN H OR CG5460.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidae; Drosophilidae; Drosophila.
 OK NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=92387549; PubMed=1516831;
 RA Bang A.G., Posakony J.W.;
 RT "The Drosophila gene Hairless encodes a novel basic protein that
 RL controls alternative cell fates in adult sensory organ development.";
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=93041287; PubMed=1419850;
 RA Maier D., Stumm G., Kuhn K., Preiss A.;
 RT "Hairless, a Drosophila gene involved in neural development, encodes
 RL a novel, serine rich protein.";
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RC STRAIN=Berkeley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Aamatiides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Chang M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abrita J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Bencos P.V., Bertman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,
 RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferlita S., Fleischmann W.,
 RA Fostel C., Gabrielian A.E., Gang N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpem G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasro P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Matrei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Modarri C., Morris J., Mostrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusseren D.R., Pacle J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stappleton M., Strong R., Sun E.,
 RA Svitak R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.W., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 CC Science 287:2185-2195 (2000).
 -1- FUNCTION: IS A POTENT ANTAGONIST OF NEUROGENIC GENE ACTIVITY

CC DURING SENSORY ORGAN DEVELOPMENT. THE EXPRESSION OF DISTINCT CELL
 CC PATES BY THE TRICHOGEN (SHAFT) / TOMOGEN (SOCKET) SISTER CELL
 CC PAIR DEPENDS ON THE LEVEL OF H ACTIVITY. A CERTAIN THRESHOLD LEVEL
 CC OF H ACTIVITY IS REQUIRED, BELOW WHICH BOTH SISTER CELLS ADOPT THE
 CC TOMOGEN FATE.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q02308-1; Sequence=displayed;
 CC Name=2;
 CC IsoId=Q02308-2; Sequence=VSP_006952;
 CC Note=No experimental confirmation available;
 CC -1- TISSUE SPECIFICITY: OVARY, EMBRYOS, LARVAL AND PUPAL IMAGINAL
 CC DISCS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M95192; AAA28607.1; ALT INIT.
 CC DR EMBL; X67239; CAA47664.1; -.
 CC DR EMBL; AE003731; AAF55790.1; -.
 CC DR EMBL; AB003731; AAF55791.1; -.
 CC PIR; A44067; A44067.
 CC DR FlyBase; FBgn0001169; H.
 CC DR GO; GO:0003714; P:transcription co-repressor activity; IDA.
 CC DR GO; GO:0007219; P:N receptor signaling pathway; NAS.
 CC DR GO; GO:0008052; P:sensory organ determination; IMP.
 CC KW Developmental protein; Nuclear protein; DNA-binding;
 CC Alternative splicing.
 CC FT DOMAIN 115 123 THR-RICH.
 CC FT DOMAIN 642 648 POLY-SER.
 CC FT DOMAIN 879 891 POLY-ALA.
 CC FT DOMAIN 937 946 POLY-ALA.
 CC FT DOMAIN 964 974 ALA-RICH.
 CC FT DOMAIN 979 1008 HIS/PRO-RICH (PRO MOTIF).
 CC FT VARSPLIC 1 18 Missing (in isoform 2).
 CC FT CONFLICT 151 151 /Ftrd=VSP_006952.
 CC FT CONFLICT 680 680 S -> A (IN REF. 2).
 CC FT CONFLICT 702 703 S -> F (IN REF. 3).
 CC FT CONFLICT 891 891 QH -> LL (IN REF. 2).
 CC FT CONFLICT 964 967 A -> R (IN REF. 2).
 CC FT CONFLICT 974 974 AAVA -> RLLP (IN REF. 2).
 CC FT CONFLICT 974 974 MISSING (IN REF. 2).
 CC SQ SEQUENCE 1077 AA; 111039 MW; A94BF1A27579E2F1 CRC64;
 CC -----
 CC Query Match 93.3%; Score 42; DB 1; Length 1077;
 CC Best Local Similarity 90.9%; Pred. No. 25;
 CC Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 CC -----
 CC Oy 1 AAAAAAAAAAAX 11
 CC Db 937 AAAAAAAAAAAR 947

OX NCBI TaxId=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92249205; PubMed=1576962;
 RA Macdonald P.M.;
 RT "The Drosophila pumilio gene: an unusually long transcription unit
 RT and an unusual protein.";
 RL Development 114:221-234(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93093466; PubMed=1459455;
 RA Barker D.D., Wang C., Moore J., Dickinson L.K., Lehmann R.;
 RT "Pumilio is essential for function but not for distribution of the
 RT Drosophila abdominal determinant Nanos.";
 RL Genes Dev. 6:2312-2326(1992).
 CC -1- FUNCTION: PUM IS THE ONLY GENE REQUIRED FOR NOS ACTIVITY THAT IS
 CC NOT ALSO REQUIRED FOR POSTERIOR LOCALIZATION OF GERM LINE
 CC DETERMINANTS. PUM IS REQUIRED DURING EMBRYOGENESIS WHEN NOS
 CC ACTIVITY APPARENTLY MOVES ANTERIORLY FROM THE POSTERIOR POLE.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. IT IS CONCENTRATED IN THE
 CC CORTICAL REGION OF THE EMBRYO BENEATH THE NUCLEI.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN THE OVARIES AND DURING THE
 CC EMBRYOGENESIS.
 CC -1- DOMAIN: CONSISTS MAINLY OF REGIONS ENRICHED IN A SINGLE AMINO
 CC ACID.
 CC -1- DISEASE: LETHAL DEFECTIVE IN POSTERIOR PATTERN FORMATION.
 CC -1- SIMILARITY: BELONGS TO THE PUMILIO/MPIS FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X62589; CAA44474.1; -.
 CC DR EMBL; L07943; AAB59189.1; -.
 CC DR PIR; A46221; A46221.
 CC DR FlyBase; FBgn0003165; pum.
 CC DR GO; GO:0008258; P:head involution; IMP.
 CC DR InterPro; IPR001313; Pumilio/Put.
 CC DR SMART; SM00025; Pumilio; 8.
 CC KW Developmental protein; Repeat.
 CC FT DOMAIN 34 45 ALA-RICH.
 CC FT DOMAIN 57 77 GLY/VNL-RICH.
 CC FT DOMAIN 83 93 ALA-RICH.
 CC FT DOMAIN 130 174 GLY-RICH.
 CC FT DOMAIN 152 164 POLY-GLY.
 CC FT DOMAIN 181 212 ALA-RICH.
 CC FT DOMAIN 213 236 GLN-RICH.
 CC FT DOMAIN 262 274 POLY-GLN.
 CC FT DOMAIN 571 599 GLY-RICH.
 CC FT DOMAIN 708 725 POLY-GLN.
 CC FT DOMAIN 936 946 POLY-ALA.
 CC FT DOMAIN 1050 1062 POLY-ALA.
 CC FT DOMAIN 1111 1326 6 X 36 AA APPROXIMATE TANDEM REPEATS.
 CC FT REPEAT 1111 1146 1.
 CC FT REPEAT 1147 1182 2.
 CC FT REPEAT 1183 1218 3.
 CC FT REPEAT 1219 1254 4.
 CC FT REPEAT 1255 1290 5.
 CC FT REPEAT 1291 1326 6.
 CC FT CONFLICT 362 362 A -> S (IN REF. 2).
 CC FT CONFLICT 1103 1103 R -> P (IN REF. 2).
 CC FT CONFLICT 1406 1407 KN -> PH (IN REF. 2).
 CC FT CONFLICT 1496 1496 V -> I (IN REF. 2).
 CC FT CONFLICT 1519 1519 S -> G (IN REF. 2).
 CC SQ SEQUENCE 1533 AA; 157521 MW; C453A2321B8BDBDC CRC64;
 CC -----
 CC Query Match 93.3%; Score 42; DB 1; Length 1533;
 CC Best Local Similarity 90.9%; Pred. No. 32;

Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAA 11
 |||||
 Db 937 AAAAAAAAAA 947

RESULT 13
 AND LIMPE
 ID ANP LIMPE STANDARD; PRT; 97 AA.

AC P09031;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 01-NOV-1988 (Rel. 09, Last sequence update)
 DT 01-AUG-1990 (Rel. 15, Last annotation update)
 DE Antifreeze protein precursor (AFP).
 OS Limanda ferruginea (Yellowtail flounder).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
 OC Pleuronectidae; Pleuronectidae; Limanda.
 OC NCBI_TaxId=8258;

RP SEQUENCE FROM N.A.
 RX MEDLINE=88029483; PubMed=3665937;
 RA Scott G.K., Davies P.L., Shears M.A., Fletcher G.L.;
 RT "Structural variations in the alanine-rich antifreeze proteins of the
 RL pleuronectinae";
 RT Eur. J. Biochem. 168:629-633(1987).
 CC -1- FUNCTION: ANTIFREEZE PROTEINS LOWER THE BLOOD FREEZING POINT.
 CC -1- SIMILARITY: BELONGS TO THE TYPE-I AFP FAMILY. TYPE 1 AFP ARE
 CC ALANINE-RICH, AMPHIPHILIC AND ALPHA-HELICAL.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC
 CC EMBL; X06356; CAA29655.1; -
 DR PIR; S02376; S02376.
 DR InterPro; IPR000104; Antifreeze_1.
 DR PRINTS; PR00308; ANTIFREEZE1.
 KM Antifreeze protein; Repeat; Signal.
 FT SIGNAL 1 23
 FT PROPEP 24 48
 FT CHAIN 49 97
 FT SEQUENCE 97 AA; 8865 MW; 62AD582DF8E459B6 CRC64;

QY 1 AAAAAAAAAA 11
 |||||
 Db 57 AAAAAAAAAA 67

RESULT 14
 ASH1 MOUSE
 ID ASH1 MOUSE STANDARD; PRT; 231 AA.

AC Q02067;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Achaete-scute homolog 1 (Mash-1).
 GN ASCL1 OR ASH1 OR MASH1 OR MASH-1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxId=10090;

RP SEQUENCE FROM N.A.
 RX MEDLINE=93144349; PubMed=8424959;
 RA del Amo F., Gendron-Magnire M., Gridley T.;
 RT "Cloning, sequencing and expression of the mouse mammalian
 RT achaete-scute homolog 1 (MASH1).";
 RL Biochim. Biophys. Acta 1171:323-327(1993).
 RL [2]

RP SEQUENCE FROM N.A.
 RX MEDLINE=94031847; PubMed=8217843;

RA Guillemot F., Joyner A.L.;
 RT "Dynamic expression of the murine Achaete-Scute homologue Mash-1 in
 RT the developing nervous system";
 RL Mech. Dev. 42:171-185(1993).

CC -1- FUNCTION: MAY PLAY A ROLE AT EARLY STAGES OF DEVELOPMENT OF
 CC SPECIFIC NEURAL LINEAGES IN MOST REGIONS OF THE CNS, AND OF
 CC SEVERAL LINEAGES IN THE PNS. ESSENTIAL FOR THE GENERATION OF
 CC OLFACTORY AND AUTONOMIC NEURONS. ACTIVATES TRANSCRIPTION BY
 CC BINDING TO THE E BOX (5'-CANNNG-3').
 CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
 CC BHLH PROTEIN. FORMS A HETERODIMER WITH E12/E47.

CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- TISSUE SPECIFICITY: DEVELOPING CNS AND PNS AT EMBRYONIC AND POST-
 CC NATAL STAGES.

CC -1- DEVELOPMENTAL STAGE: BETWEEN EMBRYONIC DAYS 8.5 AND 10.5 IT IS
 CC FOUND IN THE NEUROEPITHELIUM OF THE MIDBRAIN AND VENTRAL
 CC FOREBRAIN, AS WELL AS IN THE SPINAL CORD. BETWEEN DAYS 10.5 AND
 CC 12.5 ITS EXPRESSION PATTERN CHANGES FROM A RESTRICTED TO A
 CC WIDESPREAD ONE, IT IS THEN FOUND AT VARIABLE LEVELS IN THE
 CC VENTRICULAR ZONE IN ALL REGIONS OF THE BRAIN. FROM DAY 12.5 TO
 CC POST-NATAL STAGES IT IS ALSO EXPRESSED IN CELLS OUTSIDE OF THE
 CC VENTRICULAR ZONE THROUGH THE BRAIN, AND IN ADDITION IT IS ALSO
 CC EXPRESSED DURING DEVELOPMENT OF THE OLFACTORY EPITHELIUM AND
 CC NEURAL RETINA.

CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
 CC TRANSCRIPTION FACTORS. ASC SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; M95603; AAA37780.1; -
 DR PIR; S28186; S28186.
 DR TRANSFAC; T01619; -
 DR MGD; MGI:96919; Ascl1.
 DR GO; GO:0005634; C:nucleus; IDA.
 DR GO; GO:007400; P:neuroblast cell fate determination; IMP.
 DR InterPro; IPR001092; HLH_basic.
 DR Pfam; PF00010; HLH_1.
 DR SMART; SM00353; HLH; 1.
 DR PROSITE; PS00038; HLH_1; 1.
 DR PROSITE; PS50888; HLH_2; 1.
 KM Neurogenesis; Differentiation; Developmental protein; DNA-binding;
 KW Nuclear protein.

FT DOMAIN 30 43
 FT DOMAIN 47 52
 FT DNA BIND 116 126
 FT DOMAIN 127 166
 FT SEQUENCE 231 AA; 24755 MW; 5FAB4A62052ABF7 CRC64;

QY 1 AAAAAAAAAA 11
 |||||
 Db 34 AAAAAAAAAA 44

Query Match 91.1%; Score 41; DB 1; Length 231;
 Best Local Similarity 90.9%; Pred. No. 11;
 Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Search completed: February 19, 2004, 10:03:42
Job time : 12 secs

```
RESULT 15
ASH1_RAT STANDARD; PRT; 233 AA.
ID ASH1_RAT
AC p19359;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Achaeete-scute homolog 1.
GN ASCL1 OR ASH1 OR MASH-1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley;
RX MEDLINE=90363294; PubMed=2392153;
RA Johnson J.E., Birren S.J., Anderson D.J.;
RT "Two rat homologues of Drosophila achaeete-scute specifically
RL Nature 346:858-861(1990).
CC -1- FUNCTION: MAY PLAY A ROLE AT EARLY STAGES OF DEVELOPMENT OF
CC SPECIFIC NEURAL LINEAGES IN MOST REGIONS OF THE CNS, AND OF
CC SEVERAL LINEAGES IN THE PNS. ESSENTIAL FOR THE GENERATION OF
CC OLFACTORY AND AUTONOMIC NEURONS. ACTIVATES TRANSCRIPTION BY
CC BINDING TO THE E BOX (5'-CANNYG-3').
CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC BHLH PROTEIN. FORMS A HETERODIMER WITH E12/E47.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- TISSUE SPECIFICITY: DEVELOPING CNS AND PNS AT EMBRYONIC AND POST-
CC NATAL STAGES.
CC -1- DEVELOPMENTAL STAGE: IT IS FIRST OBSERVED AFTER NEURULATION, IN
CC 10.5 DAY OLD (E10.5) RAT EMBRYOS, AND IS RESTRICTED TO SUBSETS OF
CC NEUROEPITHELIAL CELLS IN THE SPINAL CORD AND THE BRAIN. BETWEEN
CC E10.5 AND E13.5, IN THE PERIPHERY, ITS EXPRESSION IS RESTRICTED TO
CC SOME LINEAGES OF NEURAL CREST-DERIVED CELLS, NAMELY IN SYMPATHETIC
CC AND ENTERIC NEURAL PRECURSORS. IN THE PNS ITS EXPRESSION IS
CC EXTINGUISHED AT OR BEFORE DIFFERENTIATION.
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. ASC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X53725; CAA37760.1; -
DR PIR; S11563; S11563.
DR TRANSPAC; T00484; -
DR InterPro; IPR001092; HLH_basic.
DR Pfam; PF00010; HLH_1.
DR SMART; SM00353; HLH_1.
DR PROSITE; PS00038; HLH_1; 1.
DR PROSITE; PSS0888; HLH_2; 1.
KW Neurogenesis; Differentiation; Developmental protein; DNA-binding;
KW Nuclear protein.
FT DOMAIN 30 44 POLY-ALA.
FT DOMAIN 48 54 POLY-GLN.
FT DNA BIND 118 128 BASIC DOMAIN.
FT DOMAIN 129 168 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
SQ SEQUENCE 233 AA; 24972 MW; 036BDAC8E2D3274 CRC64;

Query Match 91.1%; Score 41; DB 1; Length 233;
Best Local Similarity 90.9%; Pred. No. 11;
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAK 11
DB 35 AAAAAAAAAAQ 45
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 19, 2004, 10:01:13 / Search time 35 Seconds
(without alignments)
81.102 Million cell updates/sec

Title: US-09-551-336B-1
Perfect score: 45
Sequence: 1 AAAAAAAAAA 11

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 830525 seqs, 258052604 residues
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database:

SPTREMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp Vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	100.0	102	11	Q8C1R8
2	45	100.0	171	11	Q8CH22
3	45	100.0	396	10	Q864Y0
4	45	100.0	397	4	Q8N1Z3
5	45	100.0	594	10	Q8G8B1
6	45	100.0	602	4	Q1S337
7	45	100.0	635	4	Q1S336
8	45	100.0	687	4	Q8NB14
9	45	100.0	757	4	Q14235
10	45	100.0	804	5	Q9W2V2
11	45	100.0	1430	5	Q23590
12	45	100.0	1710	16	Q8X1G0
13	45	93.3	161	11	Q8CH17
14	42	93.3	276	11	Q8VH35
15	42	93.3	326	10	Q9FTU6
16	42	93.3			Q9FCJ6

17	42	93.3	333	10	Q9C9G2	Q9C9G2 arabidopsis
18	42	93.3	336	10	Q8S307	Q8S307 arabidopsis
19	42	93.3	336	10	Q8VZD2	Q8VZD2 arabidopsis
20	42	93.3	377	10	Q8RW2	Q8RW2 oryza sativ
21	42	93.3	399	10	Q8S608	Q8S608 oryza sativ
22	42	93.3	421	5	Q9J119	Q9J119 anthrea p
23	42	93.3	436	5	Q967T8	Q967T8 anthrea p
24	42	93.3	465	10	Q940C5	Q940C5 arabidopsis
25	42	93.3	508	11	Q9J1J6	Q9J1J6 mus musculu
26	42	93.3	619	5	Q960W5	Q960W5 drosophila
27	42	93.3	764	5	Q24708	Q24708 drosophila
28	42	93.3	814	10	Q9M2W4	Q9M2W4 arabidopsis
29	42	93.3	889	10	Q8R2J2	Q8R2J2 oryza sativ
30	42	93.3	935	5	Q81G48	Q81G48 drosophila
31	42	93.3	1077	5	Q9VDK1	Q9VDK1 drosophila
32	42	93.3	1184	5	Q9VH6	Q9VH6 drosophila
33	42	93.3	1207	5	Q9V315	Q9V315 drosophila
34	42	93.3	1533	5	Q9VH4	Q9VH4 drosophila
35	42	93.3	1607	5	Q94599	Q94599 leishmania
36	42	93.3	2175	5	Q9W198	Q9W198 drosophila
37	42	93.3	2639	5	Q76786	Q76786 anthrea p
38	42	93.3	2655	5	Q964F4	Q964F4 anthrea p
39	41	91.1	42	5	Q25210	Q25210 junonia coe
40	41	91.1	92	5	P91768	P91768 manduca sex
41	41	91.1	158	10	Q9MBF7	Q9MBF7 illium long
42	41	91.1	164	11	Q8BNH5	Q8BNH5 mus musculu
43	41	91.1	171	10	Q8H8G6	Q8H8G6 oryza sativ
44	41	91.1	217	11	Q9D801	Q9D801 mus musculu
45	41	91.1	217	11	Q9CR57	Q9CR57 mus musculu

ALIGNMENTS

RESULT 1

Q8C1R8 PRELIMINARY; PRT; 102 AA.

AC Q8C1R8; 01-MAR-2003 (TREMBLrel. 23, Created)

DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)

DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)

DT Hypothetical alanine-rich region/cysteine-rich region/EGF-like domain/type I antifreeze protein containing protein.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Testis;

RX MEDLINE=22354683; PubMed=12466651;

RA The FANTOM Consortium,

RT "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs."

RL Nature 420:563-573 (2002).

DR EMBL; AK005674; BAC25119.1; -

KW Hypothetical protein.

SQ SEQUENCE 102 AA; 11366 MW; E55722E863D42572 CRC64;

Query Match 100.0%; Score 45; DB 11; Length 102;
Best local similarity 100.0%; Pred. No. 11;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAA 11
|||||
Db 13 AAAAAAAAAA 23

RESULT 2

Q8CH22 PRELIMINARY; PRT; 171 AA.

ID Q8CH22; 08CH22;

DT 01-MAR-2003 (TReMBLrel. 23, Created)
 DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
 DE Basic protein CTR1.
 GN CTR1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN NCBI_TaxID=10090;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CD-1; Tissue=Testis; Escalier D., McFarland L., Xu P.-X.;
 RA Xu X., Bai X., Silvius D., Escalier D., McFarland L., Xu P.-X.;
 RT "Ck2 differentially phosphorylate a family of novel spermatid-specific
 basic nuclear proteins."
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF463500; AAC15673.1; -
 SQ SEQUENCE 171 AA; 19175 MW; 2482A9F50B121B68 CRC64;
 Query Match 100.0%; Score 45; DB 11; Length 171;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAA 11
 DB 13 AAAAAAAAAA 23

RESULT 3
 Q8S4Y0 PRELIMINARY; PRT; 396 AA.
 ID 08S4Y0;
 DT 01-JUN-2002 (TReMBLrel. 21, Created)
 DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
 DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
 DE Putative ERBBP-type transcription factor.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Eriactoidae; Oryzaceae; Oryza.
 RN NCBI_TaxID=4530;
 [1]
 RP SEQUENCE FROM N.A.
 RA Park M., Moon E., Hwang D.-J.;
 RT "Molecular cloning of a putative ERBBP-type transcription factor in
 rice."
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF364176; AAM00285.1; -
 DR Gramineae; Q8S4Y0; -
 DR InterPro; IPR001471; TF_ERF.
 DR Pfam; PF00847; AP2-domain; 1.
 DR PRODOM; PD001423; TF_ERF; 1.
 DR SMART; SM00380; AP2; 1.
 SQ SEQUENCE 396 AA; 42597 MW; EC3FD53BFF5F9381 CRC64;
 Query Match 100.0%; Score 45; DB 10; Length 396;
 Best Local Similarity 100.0%; Pred. No. 37;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAA 11
 DB 15 AAAAAAAAAA 25

RESULT 4
 Q8N123 PRELIMINARY; PRT; 397 AA.
 ID 08N123;
 DT 01-OCT-2002 (TReMBLrel. 22, Created)
 DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
 DE Hypothetical protein FJ37192.
 OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN NCBI_TaxID=9606;
 [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Alzheimer cortex;
 RA Kawakami B., Sugiyama A., Takemoto M., Sugiyama T., Irie R.,
 RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
 RA Kawai-Hiro Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
 RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuna M.,
 RA Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Suzuki Y.,
 RA Sugano S., Nagahara K., Masuno Y., Nagai K., Isega T.;
 RT "NEDD human cDNA sequencing project."
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: CONTAINS 2 LIM DOMAINS. THE LIM DOMAIN BINDS 2 ZINC
 IONS.
 DR EMBL; AK094511; BAC04371.1; -
 DR InterPro; IPR001356; Homeobox.
 DR InterPro; IPR001781; LIM.
 DR InterPro; IPR007107; LIM_homeo.
 DR Pfam; PF00046; homeobox; 1.
 DR Pfam; PF00412; LIM; 2.
 DR PRODOM; PD000010; Homeobox; 1.
 DR PRODOM; PD000094; LIM; 2.
 DR SMART; SM00094; LIM; 2.
 DR SMART; SM00389; HOX; 1.
 DR SMART; SM00132; LIM; 2.
 DR PROSITE; PS00027; HOMEBOX 1; 1.
 DR PROSITE; PS50071; HOMEBOX 2; 1.
 DR PROSITE; PS00478; LIM_DOMAIN 1; 2.
 DR PROSITE; PS50023; LIM_DOMAIN 2; 2.
 DR PROSITE; PS50558; LIM_HOMEDOMAIN; 1.
 DR Hypothetical protein; LIM domain; Metal-binding; Zinc.
 KM SEQUENCE 397 AA; 43363 MW; D4DB2956EA647F8D CRC64;
 Query Match 100.0%; Score 45; DB 4; Length 397;
 Best Local Similarity 100.0%; Pred. No. 37;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAA 11
 DB 178 AAAAAAAAAA 188

RESULT 5
 Q8GSB1 PRELIMINARY; PRT; 594 AA.
 ID 08GSB1;
 DT 01-MAR-2003 (TReMBLrel. 23, Created)
 DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
 DE Putative calcium/calmodulin-dependent protein kinase CamK.
 GN P0524G08.10 OR OJ1340 C08.39.
 OS Oryza sativa (Japanese cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Eriactoidae; Oryzaceae; Oryza.
 RN NCBI_TaxID=39947;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, PAC
 clone:PO524G08."
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA Sasaki T., Matsumoto T., Katayose Y.;
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, BAC
 clone:OJ1340 C08."
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AP004671; BAC16472.1; -
 DR EMBL; AP005292; BAC45222.1; -

```

KW Kinase. 594 AA; 65242 MW; 46FFFAE113CE6D7 CRC64;
SQ SEQUENCE
Query Match 100.0%; Score 45; DB 10; Length 594;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAA 11
   |||||
   |||||
Db 36 AAAAAAAAAA 46

RESULT 6
O15337 PRELIMINARY; PRT; 602 AA.
ID O15337
AC O15337;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Elastin (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Li D.Y., Toland A.E., Boak B.B., Atkinson D.L., Ensing G.J.,
RA Morris C.A., Keating M.T.;
RT "Elastin point mutations cause an obstructive vascular disease,
RT supravulvar aortic stenosis.";
RU Hum. Mol. Genet. 0:0-0(1997).
DR EMBL; U93037; AAB65621.1; JOINED.
DR EMBL; U93034; AAB65620.1; JOINED.
DR EMBL; U93035; AAB65620.1; JOINED.
DR EMBL; U93036; AAB65620.1; JOINED.
DR InterPro; IPR001179; FKBP_PPIase.
DR InterPro; IPR003979; tropoelastin.
DR PRINTS; PR01500; TROPOLASTIN.
DR PROSITE; PS00453; FKBP_PPIASE_1; 1.
FT NON TER 1
FT NON TER 602
SQ SEQUENCE 602 AA; 51807 MW; 53B5B9A71EF04807 CRC64;

Query Match 100.0%; Score 45; DB 4; Length 602;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAA 11
   |||||
   |||||
Db 274 AAAAAAAAAA 284

RESULT 7
O15336 PRELIMINARY; PRT; 635 AA.
ID O15336
AC O15336;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Elastin (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Li D.Y., Toland A.E., Boak B.B., Atkinson D.L., Ensing G.J.,
RA Morris C.A., Keating M.T.;
RT "Elastin point mutations cause an obstructive vascular disease,
RT supravulvar aortic stenosis.";
RU Hum. Mol. Genet. 0:0-0(1997).
DR EMBL; U93037; AAB65621.1; JOINED.
DR EMBL; U93034; AAB65621.1; JOINED.

```

```

DR EMBL; U93035; AAB65621.1; JOINED.
DR EMBL; U93036; AAB65621.1; JOINED.
DR InterPro; IPR001179; FKBP_PPIase.
DR InterPro; IPR003979; tropoelastin.
DR PRINTS; PRO1500; TROPOLAELASTIN.
DR PROSITE; PS00453; FKBP_PPIASE_1; 1.
FT NON_TER 1
FT TER 1
SQ SEQUENCE 635 AA; 55279 MW; 72950C364127B2A4 CRC64;

OY 1 AAAAAAAAAAAK 11
   |||||
   |||||
Db 274 AAAAAAAAAAAK 284

RESULT 8
Q8NB14 PRELIMINARY; PRT; 643 AA.
AC Q8NB14;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hypothetical protein NT2RP3J003474.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
CX NCBI_TaxId=9606;
RN (1)
RP SEQUENCE FROM N.A.
RA Ota T., Nishikawa T., Suzuki Y., Kawai-Hio Y., Hayashi K., Ishii S.,
RA Saito K., Yamamoto J., Makamatsu A., Nagai T., Nakamura Y.,
RA Nagahara K., Sugano S., Isogai T.;
RL "HI human cDNA sequencing project."
RL EMBL; AK075554; BAC11696.1; -.
DR InterPro; IPR001179; FKBP_PPIase.
DR InterPro; IPR000158; FtsZ.
DR InterPro; IPR003979; tropoelastin.
DR PRINTS; PRO0423; CELLDVISTSZ.
DR PRINTS; PRO1500; TROPOLAELASTIN.
DR PROSITE; PS00453; FKBP_PPIASE_1; 1.
KW Hypothetical protein.
SQ SEQUENCE 643 AA; 55629 MW; PFFC042617E72A69 CRC64;

Query Match 100.0%; Score 45; DB 4; Length 643;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAAK 11
   |||||
   |||||
Db 266 AAAAAAAAAAAK 276

RESULT 9
Q14Z35 PRELIMINARY; PRT; 687 AA.
AC Q14Z35;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Elastin.
GN ELN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
CX NCBI_TaxId=9606;
RN (1)
RP SEQUENCE FROM N.A.
RA MEDLINE=87274906; PubMed=3038460;

```

RA Indik Z., Yoon K., Morrow S.D., Cicilia G., Rosenbloom J.,
RA Rosenbloom J., Ornstein-Goldstein N.;
RT "Structure of the 3' region of the human elastin gene: great abundance
of Alu repetitive sequences and few coding sequences.";
RN Connect. Tissue Res. 16:197-211(1987).
[2]
RN SEQUENCE FROM N.A.
RX MEDLINE=87289668; PubMed=3039501;
RA Indik Z., Yeh H., Ornstein-Goldstein N., Sheppard P., Anderson N.,
RA Rosenbloom J.C., Peltonen L., Rosenbloom J.;
RT "Alternative splicing of human elastin mRNA indicated by sequence
analysis of cloned genomic and complementary DNA.";
RN Proc. Natl. Acad. Sci. U.S.A. 84:5680-5684(1987).
DR EMBL; M17282; AAC98393.1; JOINED.
DR EMBL; M16983; AAC98393.1; JOINED.
DR EMBL; M17265; AAC98393.1; JOINED.
DR EMBL; M17266; AAC98393.1; JOINED.
DR EMBL; M17267; AAC98393.1; JOINED.
DR EMBL; M17268; AAC98393.1; JOINED.
DR EMBL; M17271; AAC98393.1; JOINED.
DR EMBL; M17272; AAC98393.1; JOINED.
DR EMBL; M17273; AAC98393.1; JOINED.
DR EMBL; M17275; AAC98393.1; JOINED.
DR EMBL; M17276; AAC98393.1; JOINED.
DR EMBL; M17277; AAC98393.1; JOINED.
DR EMBL; M17278; AAC98393.1; JOINED.
DR EMBL; M17279; AAC98393.1; JOINED.
DR EMBL; M17281; AAC98393.1; JOINED.
DR HSSP; P50099; 1ZFU.
DR InterPro; IPR001179; FKBP_PPIase.
DR InterPro; IPR003979; tropoelastin.
DR PRINTS; PR01500; TROPOELASTIN.
DR PROSITE; PS00453; FKBP_PPIASE_1; 1.
SQ SEQUENCE 687 AA; 59529 MW; 864068C4C8B9F88F CRC64;
Query Match 100.0%; Score 45; DB 4; Length 687;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAAAAAAAAA 11
DB 302 AAAAAAAAAA 312
RESULT 10
Q14234
ID Q14234 PRELIMINARY; PRT; 757 AA.
AC Q14234;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Elastin.
GN ELN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=87274906; PubMed=3038460;
RA Indik Z., Yoon K., Morrow S.D., Cicilia G., Rosenbloom J.,
RA Rosenbloom J., Ornstein-Goldstein N.;
RT "Structure of the 3' region of the human elastin gene: great abundance
of Alu repetitive sequences and few coding sequences.";
RN Connect. Tissue Res. 16:197-211(1987).
[2]
RN SEQUENCE FROM N.A.
RX MEDLINE=87289668; PubMed=3039501;
RA Indik Z., Yeh H., Ornstein-Goldstein N., Sheppard P., Anderson N.,
RA Rosenbloom J.C., Peltonen L., Rosenbloom J.;
RT "Alternative splicing of human elastin mRNA indicated by sequence
analysis of cloned genomic and complementary DNA.";
RN Proc. Natl. Acad. Sci. U.S.A. 84:5680-5684(1987).

DR EMBL; M17282; AAC98395.1; JOINED.
DR EMBL; M16983; AAC98395.1; JOINED.
DR EMBL; M17265; AAC98395.1; JOINED.
DR EMBL; M17266; AAC98395.1; JOINED.
DR EMBL; M17267; AAC98395.1; JOINED.
DR EMBL; M17268; AAC98395.1; JOINED.
DR EMBL; M17270; AAC98395.1; JOINED.
DR EMBL; M17271; AAC98395.1; JOINED.
DR EMBL; M17272; AAC98395.1; JOINED.
DR EMBL; M17273; AAC98395.1; JOINED.
DR EMBL; M17274; AAC98395.1; JOINED.
DR EMBL; M17275; AAC98395.1; JOINED.
DR EMBL; M17276; AAC98395.1; JOINED.
DR EMBL; M17277; AAC98395.1; JOINED.
DR EMBL; M17278; AAC98395.1; JOINED.
DR EMBL; M17279; AAC98395.1; JOINED.
DR EMBL; M17280; AAC98395.1; JOINED.
DR EMBL; M17281; AAC98395.1; JOINED.
DR InterPro; IPR001179; FKBP_PPIase.
DR InterPro; IPR003979; tropoelastin.
DR PRINTS; PR01500; TROPOELASTIN.
DR PROSITE; PS00453; FKBP_PPIASE_1; 1.
SQ SEQUENCE 757 AA; 66136 MW; 23B7F5B8AF85C8 CRC64;
Query Match 100.0%; Score 45; DB 4; Length 757;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAAAAAAAAA 11
DB 302 AAAAAAAAAA 312
RESULT 11
Q9W2V2
ID Q9W2V2 PRELIMINARY; PRT; 804 AA.
AC Q9W2V2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE CG32683 protein.
GN CG32683 OR CG2881 OR CG2883 OR CG15303.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RN SEQUENCE FROM N.A.
RX STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter R.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
RA Balieu R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertan B.P., Bhandari D., Bolshakov S.,
RA Borokva D., Bocham M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burris K.C., Busam D.A., Butler C., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlike C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.B., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck C.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegum C.,
RA Jalili M., Kalush K., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Melnikov G., Milshina N.V., Mobarry C., Morris J., Mostrei A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Murry D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reiner K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Ramos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spindling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wattman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 "The genome sequence of *Drosophila melanogaster*.";
 Science 287:2185-2195(2000).
 (12) SEQUENCE FROM N.A.
 RA Celinker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Barton J., An H., Baldwin D., Banazon J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferreira S., Frise E., Galle R.F., Gary N.S., George R.A.,
 RA Gonzalez C., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
 RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Svirska R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 "Sequencing of *Drosophila melanogaster* genome.";
 Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 (13) SEQUENCE FROM N.A.
 RA Mitra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochank S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celinker S.E.,
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Seattle S.M.J., Smith E., Shu S., Smutnak F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 "Annotation of *Drosophila melanogaster* genome.";
 Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 (14) SEQUENCE FROM N.A.
 RA Adams M.D., Celinker S.E., Gibbs R.A., Rubin G.M., Venter J.C.;
 Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 (15) SEQUENCE FROM N.A.
 RA Playbase:
 RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 RA EMBL: A5003450; AAF4586.2; -.
 DR HSSP: P08168; 1CF1.
 DR PLAYBASE: FBgn0052683; CG32683.
 DR InterPro: IPR000698; Arrestin.
 DR Pfam: PF00339; arrestin_1.
 DR Pfam: PF02752; arrestin_C_1.
 DR ProDom: PD002099; Arrestin; 2.
 SQ SEQUENCE 804 AA; 87536 MW; 06244F0D8A9CB635 CRC64;

Query Match 100.0%; Score 45; DB 5; Length 804;
 Best Local Similarity 100.0%; Pred. No. 69;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAA 11
 Db 685 AAAAAAAAAA 695

RESULT 12
 Q23590 PRELIMINARY; PRT; 1430 AA.

AC Q23590;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE ZK783.4 protein.
 GN ZK783.4.
 OS *Caenorhabditis elegans*.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; *Caenorhabditis*.
 OC NCBI_TaxID=6239;
 RN (11) SEQUENCE FROM N.A.
 RP STRAIN=Bristol N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kerhaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkhen R.,
 RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterson R.,
 RA Watson A., Weinstein L., Wilkinson-Sproat J., Wohldman P.;
 "2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*.";
 Nature 368:32-38(1994).
 (12) SEQUENCE FROM N.A.
 RP STRAIN=Bristol N2;
 RC Favello A., Vaudin M.;
 RA "The sequence of *C. elegans* cosmid ZK783.";
 RT Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
 (13) SEQUENCE FROM N.A.
 RP STRAIN=Bristol N2;
 RC Watson R.;
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U13665; A02742.1; -.
 DR HSSP: Q92831; 1B91.
 DR WormPeP: ZK783.4; CE18461.
 DR InterPro: IPR001487; Bromodomain.
 DR InterPro: IPR004022; DDT dom.
 DR InterPro: IPR001739; MetH1-CpG_bind.
 DR InterPro: IPR001965; ZnF PHD.
 DR Pfam: PF04339; bromodomain; 1.
 DR Pfam: PF02791; DDT; 1.
 DR Pfam: PF01429; MBD; 1.
 DR Pfam: PF00628; PHD; 2.
 DR PRINTS: PR00503; BROMODOMAIN.
 DR SMART: SM00297; BROMO; 1.
 DR SMART: SM00571; DDT; 1.
 DR SMART: SM00391; MBD; 1.
 DR SMART: SM00249; PHD; 2.
 DR PROSITE: PS50014; BROMODOMAIN 2; 1.
 DR PROSITE: PS50016; ZF PHD 2; 1.
 SQ SEQUENCE 1430 AA; 160457 MW; 70749892B6859D70 CRC64;

Query Match 100.0%; Score 45; DB 5; Length 1430;
 Best Local Similarity 100.0%; Pred. No. 1,1e+02;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAA 11
 Db 114 AAAAAAAAAA 124

RESULT 13
 Q8XTG0 PRELIMINARY; PRT; 1710 AA.

DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
 DE Putative Rhs-related protein.
 GN RSP0151 OR RS02969.
 OS Ralstonia solanacearum (Pseudomonas solanacearum).
 OC Bacterioid megaplasmid.
 OC Bacterioid, Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Ralstoniaceae; Ralstonia.
 OX NCBI_TaxID=305;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GM11000;
 RA MEDLINE=2161879; PubMed=11823852;
 RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
 RA Aclat M., Billault A., Brottier P., Camus J.C., Catolico L.,
 RA Chandler M., Choiane N., Claudel-Renard C., Cunnac S., Demange N.,
 RA Gaspin C., Lavie M., Moisan A., Robert C., Sarrin W., Schiex T.,
 RA Siguer P., Thebaud P., Whalen M., Winkler P., Levy M.,
 RA Weissenbach J., Boucher C.A.;
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum";
 RL Nature 415:497-502(2002).
 DR EMBL: AL646076; CAD17302.1; -;
 DR InterPro: IPR005074; Peptidase_C39.
 DR InterPro: IPR006530; YD.
 DR Pfam: PF03412; Peptidase_C39; 1.
 DR TIGRPFAMs: TIGR01643; YD repeat_2x; 14.
 DR KW Plasmid; Complete proteome.
 SQ SEQUENCE 1710 AA; 187661 MW; AACDBE9646AAE29 CRC64;

Query Match 100.0%; Score 45; DB 16; Length 1710;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAK 11
 DB 1605 AAAAAAAAAAAK 1615

RESULT 14
 O8CH17 PRELIMINARY; PRT; 161 AA.
 AC O8CH17;
 DT 01-MAR-2003 (TReMBLrel. 23, Created)
 DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
 DE Helix-loop-helix protein.
 GN ID4.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Mammary gland;
 RA Shan L., Yu M., Snyderwine E.G.;
 RT "Cloning and functional analysis of rat Id4 gene."
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF468681; AA015695.1; -;
 SQ SEQUENCE 161 AA; 16621 MW; 2AB9E2D69C3909ED CRC64;

Query Match 93.3%; Score 42; DB 11; Length 161;
 Best Local Similarity 90.9%; Pred. No. 44;
 Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAK 11
 DB 39 AAAAAAAAAAAK 49

RESULT 15
 O8VH35 PRELIMINARY; PRT; 276 AA.
 AC O8VH35;
 DT 01-MAR-2002 (TReMBLrel. 20, Created)

DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
 DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
 DE HMG-box protein.
 GN SOX21.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/cJ;
 RA Uvanogho D.A.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY069926; AAL49967.1; -;
 DR InterPro: IPR000104; Antifreeze_1.
 DR InterPro: IPR000910; HMG_12_box.
 DR Pfam: PF00505; HMG_box; 1.
 DR PRINTS: PR00308; ANTI-FREEZE1.
 DR SMART: SM00398; HMG; 1.
 SQ SEQUENCE 276 AA; 28593 MW; 24A31E0FE24A9078 CRC64;

Query Match 93.3%; Score 42; DB 11; Length 276;
 Best Local Similarity 90.9%; Pred. No. 70;
 Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAK 11
 DB 137 AAAAAAAAAAAK 147

Search completed: February 19, 2004, 10:04:30
 Job time : 37 secs